Enhancing Transaction Monitoring Controls to Detect Money Laundering Using Machine Learning

By

Berkan Oztas

Department of Computing and Informatics

A thesis submitted in fulfilment of the requirements of
Bournemouth University for the degree of
Doctor of Philosophy

September 2024

Copyright Statement

Abstract

Money laundering poses a significant threat to global financial systems, with estimates suggesting that 2-5% of global Gross Domestic Product (GDP) is laundered annually. Despite the pivotal role of financial institutions in combating these illicit activities through anti-money laundering mechanisms, existing rule-based transaction monitoring systems are inefficient, most notably due to false positive rate exceeding 95%. These inefficiencies result in considerable operational costs and complicate the effective identification of genuine financial crimes.

This research aimed to address the limitations inherent in traditional transaction monitoring approaches by investigating advanced machine learning techniques to enhance the proficiency of detection. A systematic literature review forms the foundation of the study, encompassing an analysis of current transaction monitoring methods. It evaluates data pre-processing techniques, dataset characteristics, feature selection processes, and the application of various machine learning algorithms. This review highlights significant gaps in the application of advanced methods, such as deep learning and graph analysis, and underscores the lack of industry-wide collaboration along with the challenges posed by inconsistent and unavailable datasets.

In response to these challenges, the study explores the perspectives of anti-money laundering specialists through semi-structured interviews. These discussions reveal current shortcomings in transaction monitoring and the potential of machine learning, addressing the disconnect between the industry and academicians. Innovative solutions utilising anomaly detection and graph analysis to identify complex local and global patterns of transactions are identified as crucial for enhancing rule-based systems. The requirements for a successful approach, include the need for explainability, incorporating a feedback loop, and enhanced risk identification capabilities, are also thoroughly analysed.

Further, the thesis introduces SAML-D, a novel synthetic dataset created through agent-based and typology-based simulations designed to reflect the intricacies of money laundering schemes. The typologies incorporated within SAML-D were created based on insights gathered from specialists, existing datasets, and through comprehensive literature reviews. SAML-D includes diverse elements like geographic variables and high-risk payment types, providing a more robust platform for testing Anti-Money Laundering (AML) systems than currently available datasets.

Building on these insights, Tab-AML is developed. It employs a dual-masked transformer structure enhanced by a residual attention mechanism and a shared embedding approach. This model is evaluated using the SAML-D dataset against both conventional and deep learning models, such as TabTransformer and TabNet, demonstrating superior capabilities. Specifically, Tab-AML achieved a 93.01% ROC-AUC score, significantly reducing the false positive rate by 17% while maintaining a high true positive rate of 98%. The importance of model selection tailored to the task and dataset was underscored by additional tests on a real transaction monitoring dataset, where XGBoost was the best-performing model. Transformer-based models excelled at identifying suspicious behaviour in the initial monitoring stage with unprocessed transaction data, while XGBoost performed best in the later monitoring stage with pre-processed, case-structured data. These comparisons also underlined the Tab-AML model's adaptability and strengths, emphasising the benefits of the residual attention and shared embedding mechanisms where inter-transaction connections are critical.

This research advances the field of anti-money laundering by demonstrating the effectiveness of transformer-based architectures for transaction monitoring. The novel SAML-D dataset provides a robust testing platform, while Tab-AML shows superior performance in detecting suspicious activities and reducing false positive rates. Comparative analysis with other models underscores the necessity of tailored model selection while demonstrating how impactful deep learning models can be. This thesis contributes to both academic and practical domains of financial security, advocating for more intelligent and adaptive systems to meet the complex demands of the modern financial landscape.

Declaration

I, Berkan Oztas, confirm that the research contained in this thesis is my original work. The following chapters of this thesis were developed and prepared for publication:

Chapter 2

- B. Oztas, D. Cetinkaya, F. Adedoyin and M. Budka, ”Enhancing Transaction Monitoring Controls to Detect Money Laundering Using Machine Learning,” 2022 IEEE International Conference on e-Business Engineering (ICEBE), Bournemouth, United Kingdom, 2022, pp. 26-28. - *Won the Research Student Presentation Competition.*

- B. Oztas, D. Cetinkaya, F. Adedoyin, M. Budka, H. Dogan and G. Aksu, ”A Systematic Review of Machine Learning Approaches for Transaction Monitoring in Anti-Money Laundering,” - *Intended for submission to a peer-reviewed journal.*

Chapter 3

- B. Oztas, D. Cetinkaya, F. Adedoyin, M. Budka, H. Dogan and G. Aksu, ”Transaction monitoring in anti-money laundering: A qualitative analysis and points of view from industry,” Future Generation Computer Systems, vol. 159, 2024, pp. 161-171. ISSN 0167-739X.

- B. Oztas, D. Cetinkaya, F. Adedoyin, M. Budka, H. Dogan, and G. Aksu. ”Perspectives from Experts on Developing Transaction Monitoring Methods for Anti-Money Laundering,” in 2023 IEEE International Conference on e-Business Engineering (ICEBE), Sydney, Australia, 2023, pp. 39-46. - *Won Best Student Paper Award 2023.*

Chapter 4

- B. Oztas, D. Cetinkaya, F. Adedoyin, M. Budka, H. Dogan, and G. Aksu. ”Enhancing Anti-Money Laundering: Development of a Synthetic Transaction Monitoring Dataset,” in 2023 IEEE International Conference on e-Business Engineering (ICEBE), Sydney, Australia, 2023, pp. 47-54.

- The dataset developed in this study is publicly available on Kaggle and has been downloaded over 1,300 times. The SAML-D dataset can be accessed at: https://www.kaggle.com/datasets/berkanoztas/synthetic-transaction-monitoring-dataset-aml.

Chapter 5

- B. Oztas, D. Cetinkaya, F. Adedoyin, M. Budka, H. Dogan and G. Aksu, ”Tab-AML: A Transformer Based Transaction Monitoring Model for Anti-Money Laundering” in 2025 IEEE Conference on Artificial Intelligence (CAI), Santa Clara, California, USA. - *Won Best Paper Award 2025.*

Acknowledgements

The completion of this doctoral thesis would not have been possible without the support and encouragement of the academics, industry colleagues, friends, and family who have accompanied me throughout this journey.

I would like to express my deepest gratitude to my supervisors, Dr. Deniz Cetinkaya, Dr. Festus Adedoyin, and Prof. Marcin Budka. Their guidance, insightful feedback, and support have been instrumental in shaping both my research and personal development. My sincere appreciation extends to my industrial supervisor, Gokhan Aksu, and to Danske Bank for their invaluable support. Each brought unique expertise and perspectives that enhanced my work, and their dedication to excellence continually inspired me to strive for the highest standards.

I am profoundly thankful to Prof. Huseyin Dogan for his generous assistance and insightful discussions. His contributions added depth to my research and helped me navigate complex challenges with confidence.

I would also like to acknowledge all the industry professionals who participated in the semi-structured interviews. Their willingness to share their time, experiences, and knowledge was essential to the value of this research.

My heartfelt gratitude goes to my family - Kipriye, Gursel, and Emre - for their unconditional love, understanding, and encouragement. Their belief in me has been a constant source of strength, and they have been my pillars throughout this journey. I am forever grateful for them. I would like to express special thanks to Estefania for her unwavering support and understanding. Her encouragement and patience were invaluable, especially during the most challenging phases of this journey. Her presence provided balance and perspective, allowing me to continue with renewed determination.

Lastly, I wish to thank all my friends who have supported me along the way. Their company and shared experiences not only enriched my personal life but also kept me going throughout this journey.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and Overview

Money laundering signifies a major challenge and risk in the global economic landscape. The process involves transforming income generated through crime into money or assets that appear legitimate, enabling the criminals to utilise their proceeds (Michael Levi and Reuter, 2006). The complexities of money laundering methods have evolved with the prompt development of financial systems and technologies, making it an active threat. The United Nations Office on Drugs and Crime (UNODC) estimated that 2-5% of global GDP is laundered globally every year, corresponding to up to $2 trillion (United Nations Office on Drugs and Crime, 2021). These estimates support the danger and significance that money laundering possesses on financial institutions and governments worldwide. Although these figures are helpful, quantifying the precise amount of laundered money is difficult due to the domain's secretive nature (Butgereit, 2021).

The world is significantly impacted economically and socially by the effects of money laundering. The introduction of illicit funds into legitimate financial systems weakens the integrity of financial markets by distorting asset and commodity prices as well as reducing economic stability (Hendriyetty and Grewal, 2017). The proceeds of money laundering also facilitate crime and terrorism as it delivers economic support to such organisations, negatively affecting society (De Koker, 2006). Failing to prevent criminals from exploiting banking systems and institutions can also erode public trust as they become aware of the illegal activities it allows. Hence, attempts at reducing money laundering are crucial to prevent criminals from profiting from their activities. As well as protecting the economic landscape, reducing money laundering can decrease societal harm arising from human trafficking, drug trafficking, terrorism, and corruption (McDowell and Novis, 2001).

In this context, it is essential to distinguish between money laundering, terrorist financing, and other illicit activities, as each presents unique challenges for financial institutions and regulatory bodies. Money laundering involves legitimising the proceeds of criminal activities such as fraud, drug trafficking, and corruption (FATF, 2023). In contrast, terrorist financing refers to the provision of funds for terrorist activities, which may originate from both legal and illegal sources (FATF, 2023). Unlike money laundering, the primary objective of terrorist financing is not to conceal the source of funds but to ensure they reach terrorist organisations undetected. Additionally, other activities, such as tax evasion, fraud, and sanctions evasion, may still be flagged by transaction monitoring systems due to suspicious patterns, even though they are distinct from both money laundering and terrorist financing. Recognising these distinctions is crucial for the development of targeted monitoring and prevention strategies.

International organisations such as the Financial Action Task Force (FATF) were created to combat money laundering by providing recommendations to serve as the international standards to over 200 countries and jurisdictions (FATF, 2023). Countries worldwide have implemented the recommended legal, regulatory, and operational standards, with many establishing specialised bodies to enforce and monitor the AML regimes. Financial institutions are required to deal with regulatory pressures which can lead to legal fines and reputational damage if they have not followed the recommendation. The regulatory framework was modified from a rule-based approach to a risk-based in 2012 when FATF realised the inefficacy and ineffectiveness of the rule-based approaches (Helmy et al., 2016). Currently, FATF emphasises implementing preventive measures across financial and specific non-financial domains, enhancing legal entity transparency, assigning clear authority to relevant bodies, and promoting international cooperation to identify relevant risks.

Financial institutions conduct AML activities to fight against money laundering by combining the use of the regulatory framework and innovative technologies. Initially, financial institutions utilise foundational activities to validate the identities of their clients and generate in-depth risk profiles. Procedures such as Customer Due Diligence (CDD) and Know Your Customer (KYC) are utilised (Pieth and Aiolfi, 2003). Transaction monitoring is a further requirement for the institutions to ensure they take a proactive approach in detecting and preventing the flow of criminal funds (Simpson, 2018). Figure 1.1 gives an overview of the AML process within banks, with a focus on transaction monitoring. The transactions start by going through the banks' system where the suspicious transactions are alerted and sent to the investigation team, who assess if the alert is truly suspicious. Once it is confirmed to be a suspicious case, a suspicious activity report (SAR) is prepared and sent to the authorities. If the transaction is a false alert, the investigation ends. The authorities then analyse the case and decide if any further action is necessary.

The goal of transaction monitoring is to identify and prevent potential threats by triggering them to be investigated further. Financial institutions primarily use rule-based transaction monitoring systems and the reliance on this type of traditional technology has limitations (Z. Chen, Van Khoa, Teoh, et al., 2018). A major challenge is the number of false positive alerts created by the systems, increasing operational costs and inefficient use of resources which could be beneficial in other areas of the business (B. Oztas et al., 2023b). Due to the limitations and challenges of the rule-based methods, there is a need for innovative solutions. Therefore, the application of machine learning for transaction monitoring is an active research area aimed at addressing these challenges. Chapter 2 presents a systematic literature review on the detection of money laundering through machine learning models, specifically focusing on transaction monitoring. This specific approach to review distinguishes it from broader reviews that cover various AML strategies in different domains. The chapter presents an in-depth analysis of the data pre-processing, feature engineering, databases, algorithms, and evaluation techniques that the current solutions use. Existing work is classified and compared on the basis of the adopted machine learning methods. The shortcomings of the existing approaches and research gaps are identified at the end of the chapter. The findings highlight the dominance of anomaly

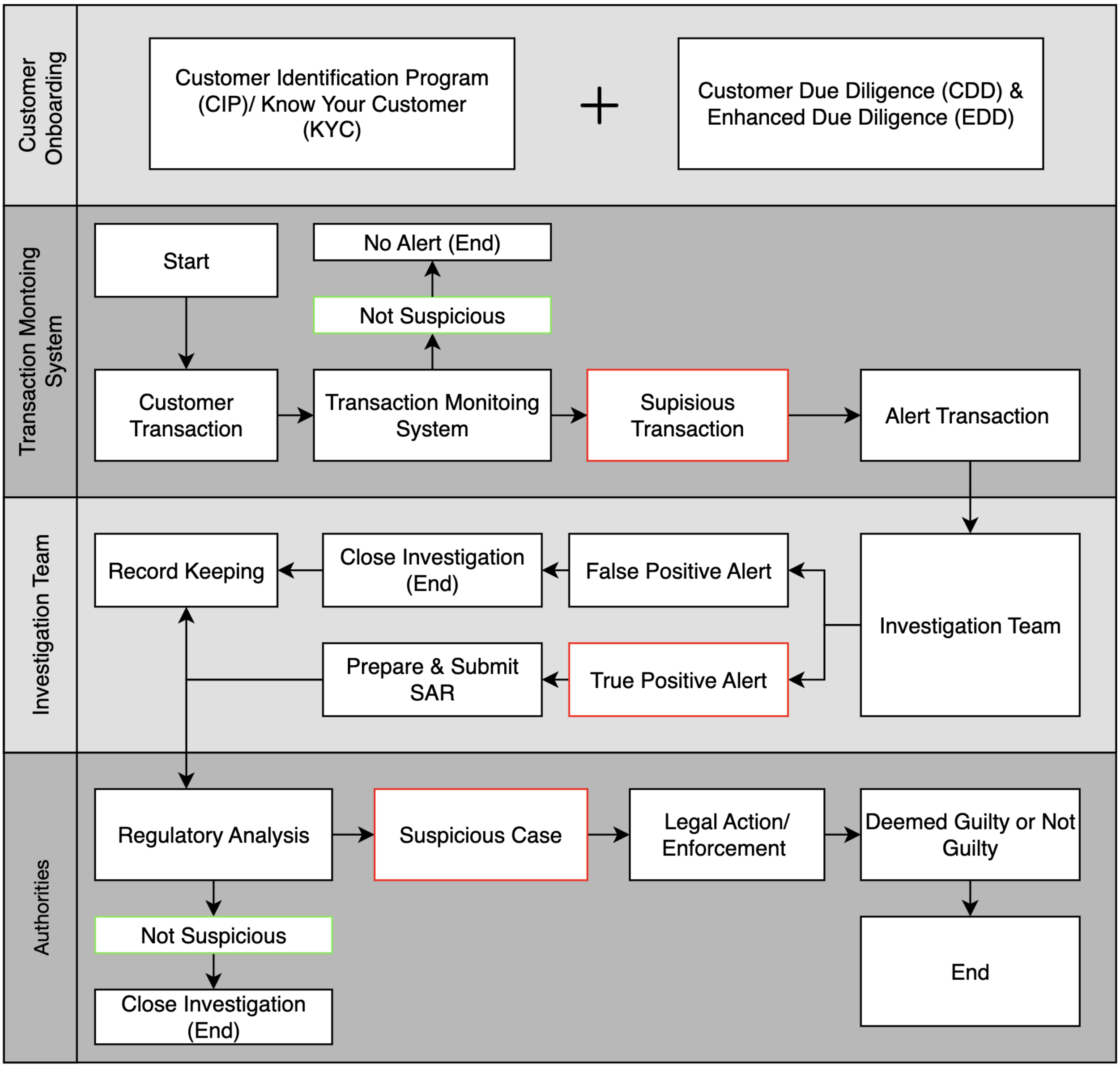


Figure 1.1: The process of AML in a financial institution. Adapted from Kute et al., 2021.

detection techniques, the need for comprehensive datasets for performance comparison, and the challenges in accessing data and collaborating with financial institutions. Additionally, it emphasises the limited research on advanced machine learning areas such as deep learning and graph machine learning. Overall, this chapter aims to assist in advancing the knowledge and support decision-making in the transaction monitoring domain.

Chapter 3 explores the perspectives of anti-money laundering specialists on the current and future state of transaction monitoring, motivated by the previously identified gap between academic and industrial collaboration. Through semi-structured interviews, this chapter aims to aid in the development of transaction monitoring approaches that are enhanced and more suitable for the industry. Three core goals are addressed: Identifying the challenges financial institutions face within the transaction monitoring department, such as data collection and quality issues, high false positive rates, significant operational costs, and rule creation; understanding the requirements for the successful adoption of new transaction monitoring approaches that allow scalability with greater accuracy, efficiency, and flexibility; and exploring the specialists' views on the potential solutions and the future direction of transaction monitoring. The chapter reveals a growing interest in artificial intelligence and machine learning to transform transaction monitoring. Specialists anticipate features such as the capability to adjust to evolving customer behaviours, uncover hidden relationships, and ensure the explainability of their results in new solutions. These results can aid researchers and stakeholders in gaining a comprehensive understanding of the challenges within the transaction monitoring sector, offering a road map for developing effective and efficient transaction monitoring strategies that align with industry needs.

Building onto the identified challenges around data in Chapters 2 and 3, a synthetic AML transaction data generator is described in Chapter 4, which outputs the SAML-D dataset. The dataset includes a range of 'normal' and 'suspicious' transactions between various agents (bank accounts) and is structured in a tabular format. SAML-D contains 12 features and 28 typologies developed from analysing existing datasets, the current literature, and insights attained from interviews with AML specialists. To generate the transactions two approaches were used: the agent-based and typology-based methods. These methods allowed for interactions between accounts and ensured each typology was significantly represented in the dataset. In comparing the SAML-D to existing synthetic datasets such as AMLSim (Suzumura and Kanezashi, 2021), IT-AML (Altman et al., 2023), and MLDP (Mahootiha, 2020) a notable improvement was the addition of geographic locations, number of typologies, and the focus put on payment types. To demonstrate a practical application of this study, experiments were conducted comparing the results of machine learning models on the SAML-D dataset and the established AMLSim datasets. Results indicated SAML-D's higher complexity over the AMLSim dataset. The fundamental aim of this chapter is to offer researchers an extra resource for assessing their models and enabling an analysis that allows for comparison of their findings. This research is motivated by the scarcity and inaccessibility of real money laundering transaction data. Even when real data is available, obtaining accurate ground truths is challenging as money laundering transactions are inherently rare and conducted in a secretive manner. Overall, this chapter addresses the challenges of accessing money laundering data.

The industry's traditional reliance on rule-based approaches is inefficient, prompting a shift towards utilising machine learning for this task. Chapter 5 introduces a deep-learning approach specifically designed to improve transaction monitoring. The chapter highlights two main contributions: the evaluation of existing unexplored deep learning models and the development of a transformer model for monitoring transactions, called Tab-AML. The Tab-AML model employs dual-masked transformer encoders with a residual attention mechanism to increase performance. To evaluate its performance, Tab-AML was compared with various well-known baseline models in the current literature using two datasets: a synthetic dataset created during this project (SAML-D), and an anonymised real dataset from an international banking institution. The comparative analysis demonstrated that the Tab-AML model outperforms the alternative models using SAML-D, by attaining a higher receiver operating characteristic (ROC) curve and ROC-AUC (area under the curve) score performance. Additionally, the model demonstrates a reduction in the False Positive Rate (FPR) and an increase in the True Positive Rate (TPR), combined with a refined clarity in the confusion matrix outcomes. The experiments on the real dataset distinguished XGBoost as the most effective model, highlighting the critical role of matching the capabilities of the model with the unique features and demands of the specific task and dataset. Chapter 5 demonstrates the effectiveness of transformers in identifying money laundering behaviour and contributes a transformer-based approach, highlighting the unexplored potential of transformers in this field. Chapter 6 continues the assessment of the deep learning models utilising the anonymised real dataset from the international banking institution and expands its use to a different transaction monitoring task. It determines the most effective scenarios for applying specific models based on the tasks and datasets, thereby refining strategies within transaction monitoring. The results highlight the benefits that transformer-based models and the residual attention mechanism offer in this field. This research is intended to encourage future academic research into the application of transformer models in detecting money laundering behaviours.

The concluding remarks of Chapter 7 summarise the contributions of this study, placing their findings within a broader perspective, and highlights the limitations. Recommendations for future research are offered, guiding future efforts to build on top of this research.

1.2 Research Aims & Objectives

1.2.1 Research Aim

This research aims to investigate advanced machine learning and deep learning techniques, in improving the detection of money laundering activities, specifically transaction monitoring. This study seeks to address the limitations of traditional rule-based transaction monitoring systems by exploring innovative approaches that can enhance efficiency, and reduce false positives.

1.2.2 Research Objectives

(i) **Comprehensively review and critically evaluate the current state of transaction monitoring technologies**. This will involve cataloguing existing machine learning models, assessing their implementation challenges, and evaluating their effectiveness in detecting money laundering activities. Achieving this objective will enhance the understanding of the strengths and limitations of these technologies, guiding the direction and focus of this research project.

(ii) **Analyse the insights and expectations of anti-money laundering specialists** regarding the challenges faced by financial institutions in AML, the requirements for adopting new technologies, and the potential of advanced machine learning approaches to meet industry needs. This objective leverages expert opinions to help AML researchers and practitioners develop more effective transaction monitoring methods.

(iii) **To develop and assess a synthetic AML transaction data generator**, aimed at overcoming the challenges related to the scarcity and sensitivity of transaction data for research. The development of this dataset will focus on the diversity in transaction typologies to enable robust testing of machine learning models.

(iv) **Develop and evaluate a transformer based approach for transaction monitoring detection**. This objective includes rigorous testing of the model's performance against various benchmarks, including traditional machine learning and deep-learning approaches. The evaluation will focus on the reduction in false positive rates, the increase in true positive rates, and the ability to detect complex laundering activities. achieving this will provide insights into the efficacy of transformer models in strengthening transaction monitoring.

(v) **Further evaluate the developed model using a real dataset and at a later phase in the transaction monitoring process**. This involves assessing the model's effectiveness in real-world settings and examining its adaptability to an advanced phase of transaction monitoring. Also, to compare the performance outcomes between the synthetic and real datasets to highlight the characteristics of the different models.

1.3 Methodology

1.3.1 Design Science

The overall methodology of this thesis is design science (DS), a problem-solving paradigm (Hevner et al., 2004) appropriate for developing new and advanced approaches to the complex challenges of money laundering detection. The DS approach is essential for gaining a comprehensive understanding of the problems, as well as for developing and evaluating a solution that meets the required standards for transaction monitoring. Figure 1.2 presents an IDEF0 diagram to visually represent the process through the five phases in this research. These phases are categorised as explicate problem, define requirements, design and development, demonstration, and evaluation (Johannesson and Perjons, 2014). Although DS provides the overarching methodology, each chapter of the thesis contains a specific methodology regarding the specific methods it utilises. In this section, I provide a concise description of the methodologies employed in each phase, to prevent repetition for the reader.

1.3.2 Problem Explication

In the initial explicate problem phase, the aim is to identify challenges in money laundering, specifically transaction monitoring, and to justify the creation and development of the artefact. This entails conducting a systematic literature review. This review process will define research questions, establish a search strategy, and specify inclusion and exclusion criteria for selecting relevant studies. Databases will be chosen for their relevance and coverage of money laundering topics. Key search terms will be identified to focus on machine learning transaction monitoring methods. A multi-staged screening process will then be employed, culminating in a data extraction stage that will systematically analyse the selected studies, emphasising machine learning methods, datasets, and evaluation metrics.

1.3.3 Define Requirements

Building upon the previous phase a qualitative approach will be employed, focusing on semi-structured interviews with AML specialists to enhance understanding of challenges and requirements in the transaction monitoring domain. This approach is chosen to capture complex

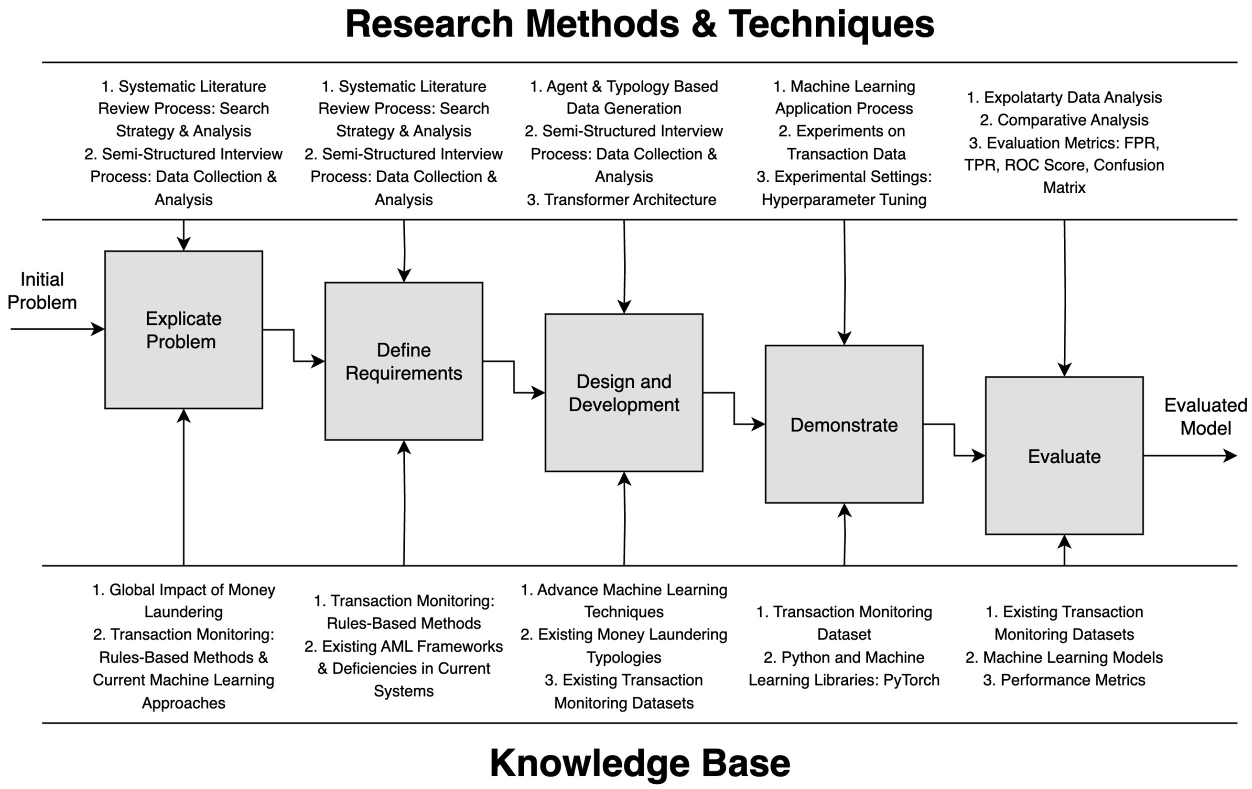

Research Methods & Techniques
1. Systematic Literature Review Process: Search Strategy & Analysis
2. Semi-Structured Interview Process: Data Collection & Analysis
1. Systematic Literature Review Process: Search Strategy & Analysis
2. Semi-Structured Interview Process: Data Collection & Analysis
1. Agent & Typology Based Data Generation
2. Semi-Structured Interview Process: Data Collection & Analysis
3. Transformer Architecture
1. Machine Learning Application Process
2. Experiments on Transaction Data
3. Experimental Settings: Hyperparameter Tuning
1. Expolatarty Data Analysis
2. Comparative Analysis
3. Evaluation Metrics: FPR, TPR, ROC Score, Confusion Matrix
Initial Problem
Explicate Problem
Define Requirements
Design and Development
Demonstrate
Evaluate
Evaluated Model
1. Global Impact of Money Laundering
2. Transaction Monitoring: Rules-Based Methods & Current Machine Learning Approaches
1. Transaction Monitoring: Rules-Based Methods
2. Existing AML Frameworks & Deficiencies in Current Systems
1. Advance Machine Learning Techniques
2. Existing Money Laundering Typologies
3. Existing Transaction Monitoring Datasets
1. Transaction Monitoring Dataset
2. Python and Machine Learning Libraries: PyTorch
1. Existing Transaction Monitoring Datasets
2. Machine Learning Models
3. Performance Metrics
Knowledge Base

Figure 1.2: Design Science Methodology Overview

details, opinions, and experiences of the specialists. Purposive sampling will be used to ensure relevant participant contributions. Data will be collected during semi-structured interviews, allowing flexibility to examine participant knowledge and potential emerging trends. An inductive approach will be adopted to analyse the data, followed by a systematic process of coding and developing themes to reflect specialist perspectives.

1.3.4 Design and Development

Building on the findings from the initial two phases, the aim will be to develop two key artefacts: a synthetic money laundering dataset and a deep-learning transaction monitoring method. An agent-based (S. H. Chen and Venkatachalam, 2017) and typology-based (Valbuena, Verburg, and Bregt, 2008) approach will be adopted to create the dataset. The first stage will focus on generating normal transaction patterns by assigning various typologies to normal bank accounts. A typology-based approach will then be employed to construct suspicious accounts that perform normal and suspicious transactions. This approach will produce a dataset with varying but overlapping transaction patterns. Selecting typologies for the dataset will involve an analysis of the existing literature, available datasets, and insights obtained from semi-structured interviews with AML specialists.

The deep-learning transaction monitoring method will be designed in three key stages: embedding, transformer encoding, and classification. The embedding stage will process the data, transforming it into a suitable format for the model to interpret the various features. The transformer stage will include two encoder blocks: the first will focus on the transaction flow to capture complex patterns, and the second will take a broader view by considering all features.

The final stage will employ a Multi-Layer Perceptron (MLP) with a sigmoid activation function for classification. This approach is chosen to address the complexity of money laundering, which often involves intricate relationships across many features. By combining sequential and feature-based analyses, it could provide a comprehensive view of the data while maintaining flexibility in handling diverse transaction scenarios.

1.3.5 Demonstrate and Evaluate

A demonstration of the synthetic transaction monitoring data will include an in-depth description of its structure, including the money laundering typologies it simulates. To validate the practicality of SAML-D, well-established machine learning models will be employed to examine the dataset's suitability for AML purposes. Exploratory data analysis will be conducted to review the dataset's distribution and alignment with expectations. A comparative analysis with existing synthetic datasets in current literature will assess coverage and relevance. Key performance metrics such as true positive rate (TPR), false positive rate (FPR), and ROC-AUC scores will be considered to gauge overall performance.

To demonstrate the transformer-based transaction monitoring model, synthetic and real datasets will be utilised to evaluate the model. Experimental settings, including the train-validation-test split and hyperparameter tuning, will be defined to optimise the model's learning and generalisation. The method will be applied to the datasets for performance evaluation, with comparisons made against transformer-based and established machine learning approaches. Metrics such as the ROC curve, ROC-AUC score, confusion matrix, FPR, and TPR will be used to assess performance, focusing on standard AML transaction monitoring considerations. This approach will allow thorough testing on diverse datasets and facilitate comparisons with multiple models.

It should be noted that the ROC-AUC score has inherent limitations. The averaging process is fundamentally incoherent because it relies on a classifier-dependent weighting scheme, which may lead to inconsistent and unfair comparisons between models (D. Hand, 2009). Consequently, the ROC-AUC score is not employed as the sole evaluation metric in this study, instead, it is supplemented with complementary metrics to ensure a more robust assessment of model performance.

1.4 Contributions

Building upon the detailed introduction, this research makes several significant contributions to the field of AML efforts within financial institutions.

Contribution 1: Bridging the gap between academic research and practical AML needs through interviews with domain specialists. Through semi-structured interviews with domain specialists in anti-money laundering, the study captures and integrates practical insights and challenges directly from the field. These interviews supplement the academic research with real-world perspectives on what is needed for effective transaction monitoring, ensuring that the research outcomes are not only theoretically sound but also practically relevant and applicable. Insights on solutions and the future of transaction monitoring are also evaluated. This contribution enhances the relevance and applicability of research findings, aiming to make them academically robust and directly applicable in real-world problems.

Contribution 2: Development of the synthetic AML transaction data generator, SAML-D. Addressing the challenge of data availability in AML research, SAML-D was developed to generate synthetic transaction data. SAML-D extends the range of normal and suspicious transaction typologies but also incorporates additional features like geographic locations and involvement of high-risk countries, adding layers of complexity compared to existing AML data. The new typologies within the dataset were developed by insights from specialists, ensuring that it closely mirrored real-world transaction challenges. Overall, SAML-D extends beyond existing AML datasets, providing a platform for the development and evaluation of new detection approaches.

Contribution 3: The application of Tab-Transformer, TabNet, and development of the Tab-AML model for transaction monitoring. The under explored application of transformers in the AML domain led to the development of Tab-AML, a transformer-based transaction monitoring model. Tab-AML builds upon existing models by integrating dual-masked attention, shared embedding, and residual attention mechanisms. The Tab-AML, Tab-Transformer, and TabNet models have been evaluated against traditional methods using the SAML-D dataset, demonstrating their effectiveness in detecting money laundering activities. Tab-AML stands out by significantly reducing false positives while maintaining accuracy, providing a valuable tool for improving AML strategies within financial institutions. This contribution highlights the potential of transformer-based deep learning models, particularly the Tab-AML model, in AML transaction monitoring.

Contribution 4: Utilisation of a real dataset to further evaluate the deep learning and baseline models for AML. The evaluations conducted on the real datasets highlight the advantages and limitations of each model when applied to specific transaction monitoring tasks. These insights confirm the importance of matching model capabilities with task demands and dataset characteristics to ensure effective detection. Additionally, this study applies these models to later stages of the transaction monitoring process, beyond the initial screening of unprocessed transactions, a novel approach not previously conducted. This advances the understanding and capabilities in detecting money laundering activities using the transformer-based models, providing practical insights that can improve AML strategies.

Overall, this research contributes to advancing the knowledge and capabilities in the detection of money laundering, offering practical insights, methodologies, and tools that can significantly enhance the effectiveness of AML strategies in the financial sector.

1.5 Organisation of the Thesis

This thesis adopts an integrated format chosen to facilitate an immediate impact within the AML field, with each chapter designed to address a specific aspect of the broader research aim. The introductory chapter in this thesis is streamlined to avoid repetition of the literature and focuses on outlining the thesis structure while linking the chapters together. This approach

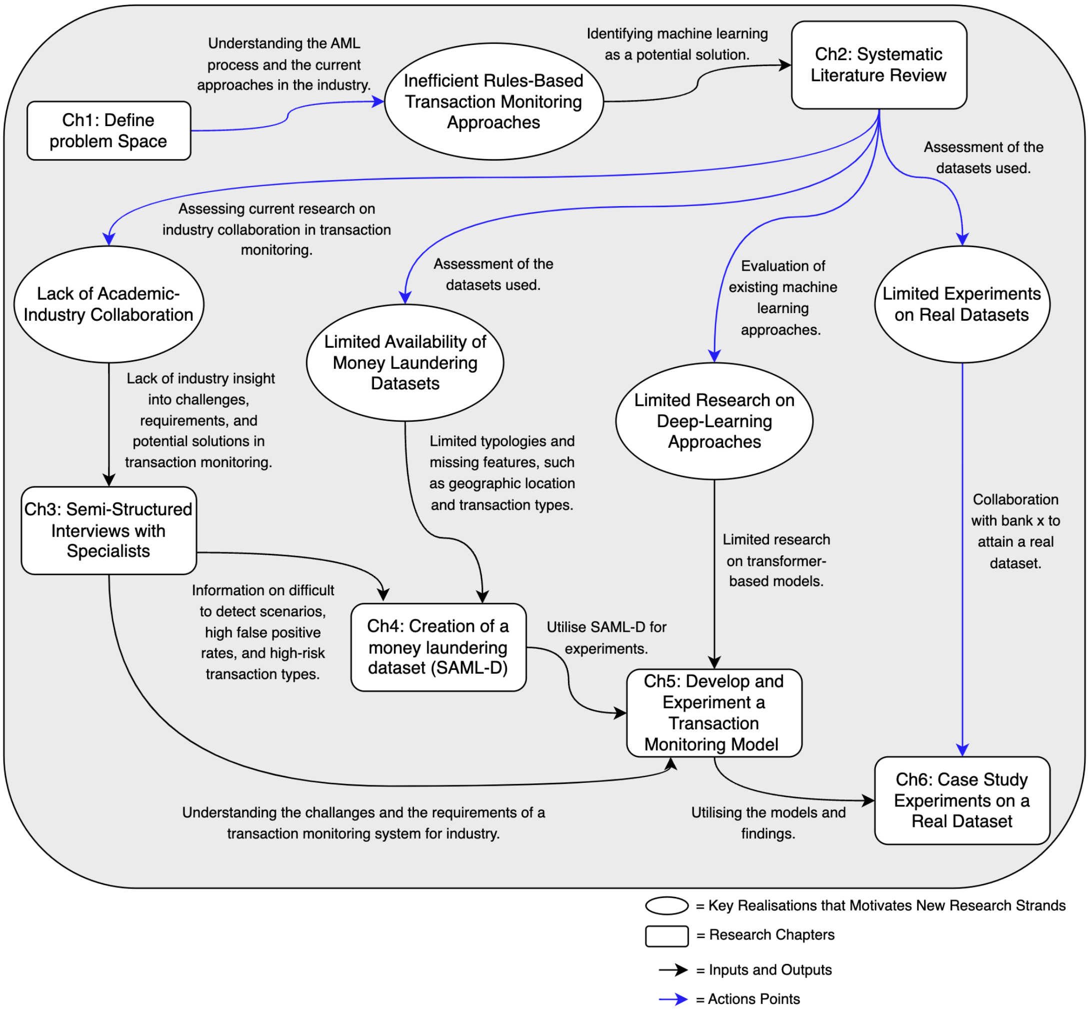

Ch1: Define problem Space
Understanding the AML process and the current approaches in the industry.
Inefficient Rules-Based Transaction Monitoring Approaches
Identifying machine learning as a potential solution.
Ch2: Systematic Literature Review
Assessment of the datasets used.
Assessing current research on industry collaboration in transaction monitoring.
Assessment of the datasets used.
Evaluation of existing machine learning approaches.
Lack of Academic-Industry Collaboration
Limited Availability of Money Laundering Datasets
Limited Research on Deep-Learning Approaches
Limited Experiments on Real Datasets
Lack of industry insight into challenges, requirements, and potential solutions in transaction monitoring.
Limited typologies and missing features, such as geographic location and transaction types.
Limited research on transformer-based models.
Collaboration with bank x to attain a real dataset.
Ch3: Semi-Structured Interviews with Specialists
Information on difficult to detect scenarios, high false positive rates, and high-risk transaction types.
Ch4: Creation of a money laundering dataset (SAML-D)
Utilise SAML-D for experiments.
Ch5: Develop and Experiment a Transaction Monitoring Model
Ch6: Case Study Experiments on a Real Dataset
Understanding the challanges and the requirements of a transaction monitoring system for industry.
Utilising the models and findings.
= Key Realisations that Motivates New Research Strands
= Research Chapters
= Inputs and Outputs
= Actions Points

Figure 1.3: Overview of the PhD

guides readers through the research journey. Chapter 2, a submitted manuscript, currently under review provides a systematic literature review on machine learning approaches for transaction monitoring. It sets a foundation for the thesis by identifying the shortcomings and gaps within the literature. Chapters 3, 4, and, 5 delve into the identified gaps in the literature, incorporating insights from AML specialists, the creation of a synthetic transaction dataset, and deep-learning methods to improve transaction monitoring. Chapter 6 expands the analysis of the models by testing them on a real dataset, exploring the impacts and strengths of the model's features in practice. Each chapter structured as a research paper includes a detailed introduction, related work, and an in-depth methodology. The concluding chapter summarises the research findings, discussing their implications in the transaction monitoring domain within a larger context. The study's limitations are acknowledged, and future research directions are suggested.

Chapter 2

Systematic Literature Review

Version of record available at: B. Oztas, D. Cetinkaya, F. Adedoyin and M. Budka, ”Enhancing Transaction Monitoring Controls to Detect Money Laundering Using Machine Learning,” 2022 IEEE International Conference on e-Business Engineering (ICEBE), Bournemouth, United Kingdom, 2022, pp. 26-28. - Won the Research Student Presentation Competition.
DOI: 10.1109/ICEBE55470.2022.00014

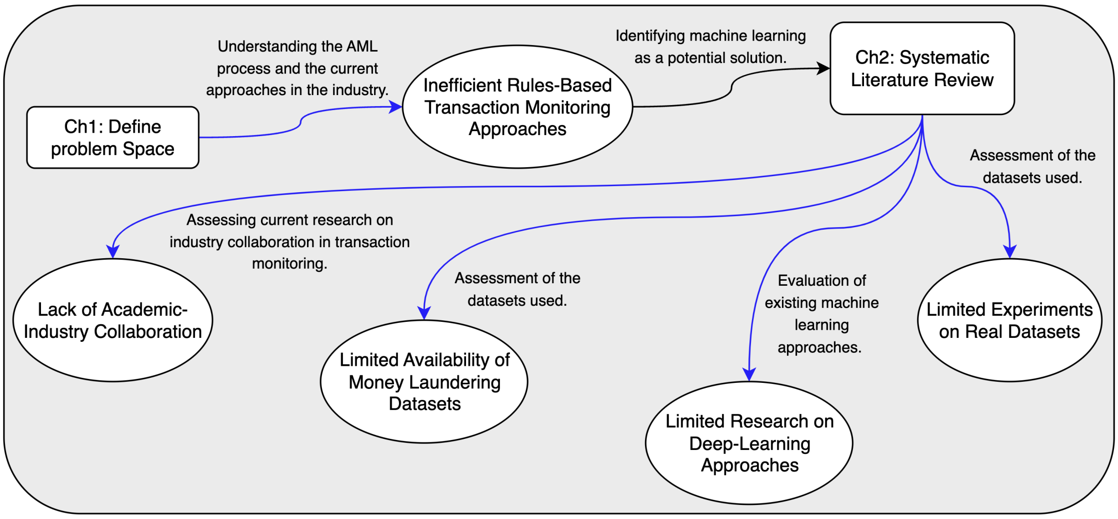

Ch1: Define problem Space
Understanding the AML process and the current approaches in the industry.
Inefficient Rules-Based Transaction Monitoring Approaches
Identifying machine learning as a potential solution.
Ch2: Systematic Literature Review
Assessment of the datasets used.
Assessing current research on industry collaboration in transaction monitoring.
Lack of Academic-Industry Collaboration
Assessment of the datasets used.
Limited Availability of Money Laundering Datasets
Evaluation of existing machine learning approaches.
Limited Research on Deep-Learning Approaches
Limited Experiments on Real Datasets

Figure 2.1: Overview of the PhD (extracted from Figure 1.3)

2.1 Introduction

This chapter builds on Chapter 1 by examining AML strategies, especially focusing on machine learning models for transaction monitoring. As financial crimes become more sophisticated, understanding how technology can help in detection is crucial. Figure 2.1 shows how this chapter fits into the overall thesis and highlights the key insights that led to new research ideas. This chapter addresses objective (i): to review and critically assess the current state of transaction monitoring technologies. This involves carrying out a systematic literature review to identify current machine learning models, examine their implementation challenges, and assess their effectiveness in detecting suspicious activities. The findings here aim to advance the transaction monitoring field and guide the broader research project. As part of this integrated thesis, this chapter has been published as a short paper to the *2022 IEEE International Conference on e-Business Engineerin*, and won the research student presentation award. Additionally, the chapter is intended for submission to a peer-reviewed journal.

Money laundering is a great economic problem with huge consequences for the society and financial institutions (McDowell and Novis, 2001). It is estimated that 2-5% of global GDP ($400bn - $2.85tn) is laundered money, however, it is difficult to gauge due to its secretive kind (Butgereit, 2021). AML aims to reduce money launderers converting criminal incomes into legal assets by referring to laws, techniques, and regulations. Currently, the risk-based framework (Helmy et al., 2016) requires financial institutions to focus primarily on customer due diligence and transaction monitoring (Naheem, 2016). The legal system has been utilised by regulators to enforce financial institutions to prioritise AML within their companies to reduce money laundering. Large fines have been issued to financial institutions all over the world that failed to meet AML standards (Jolly, 2020).

Transaction monitoring is conducted by financial institutes as they are required to continue risk assessment by monitoring patterns and anomalies in transactions of clients. Alerts are created in various systems by different actions, for example, strange transaction size, location, changes in client behaviour or unrealistic wealth in contrast with customer reports. SAR's are filed and sent to Financial Intelligence Units (FIU) when an institution considers a customer's actions suspicious.

Rule-based techniques are utilised by most financial institutions to detect transactions and flag suspicious activities. Rules are set by domain experts. The transactions are filed as SARs to FIUs. The false positive rate of the filed transactions is projected to be over 95% (Eifrem, 2019), hence being costly and time-consuming for financial institutions. There has been a huge increase in transactions over the years as financial institutions have used new technology to make services available to customers easier (i.e. through online banking). Rule-based methods have struggled to keep up with the increase in transactions and changes in customer behaviours giving inefficient results (Kute et al., 2021). With large fines being issued and the complexity of identifying financial crime, financial institutions' task of transaction monitoring is more expensive and difficult than ever.

Due to the difficulties with rule-based methods, more attention has been given to machine learning to reduce the number of false positives and acquire more efficient results (Canhoto, 2021). Machine learning approaches can also help upkeep the institution's reputation, reduce operational costs significantly, and satisfy the regulator's requirements. When applying algorithms to identify money laundering it is crucial to recognise the benefits and drawbacks of algorithms, as the effectiveness of the method is largely influenced by the unique features of the transactional data. Therefore, an overview of multiple research efforts will help in choosing appropriate algorithms and developing better approaches. Many different approaches and machine learning algorithms have been used to detect money laundering. Mainly supervised and unsupervised machine learning methods have been developed. Supervised approaches are easier to evaluate due to having a labelled dataset, however, high-quality labels are required. Unlike supervised methods, unsupervised methods can better adapt to changes in customers' behaviour and detect suspicious activity (Z. Chen, Van Khoa, Teoh, et al., 2018). However, evaluating the results of unsupervised methods is difficult as expert analysis can be needed. Instant-based approaches utilising algorithms such as support vector machine and k-means have been developed, as well as, isolation forest, local outlier factor, and, neural networks used for outlier detection. Although, deep learning techniques such as autoencoders and self-organising maps have been explored more research is required.

This systematic literature review aims to compile and analyse machine learning methods produced for transactional monitoring in anti-money laundering that use transactional data. To accomplish the aim, the following research question is formulated: "What methods have been used in the different phases of the machine learning architecture that have been developed for transactional monitoring to prevent money laundering?".

This review chapter concentrates specifically on the application of machine learning techniques in the field of transaction monitoring, as opposed to prior review articles which take a more generalised approach, encompassing multiple solutions in anti-money laundering such as KYC and screening. The key contribution of this review paper is that it conducts an analysis of the machine learning process for transaction monitoring, including data pre-processing, feature engineering, the database, the machine learning algorithms, the approach types, and the evaluation techniques used. The existing work is classified into different approaches based on the characteristics of the machine learning methods and compared. Additionally, the paper assesses and gives an evaluation level to each of the existing machine learning methods in the literature, and provides key shortcomings and future research directions in the machine learning transaction monitoring domain by examining and investigating the existing work. Findings can be used as an evidence-based guide to select appropriate approaches and algorithms, to address the identified shortcomings and future research directions. Overall, this chapter can be used for advancing knowledge and guiding decision making in the transaction monitoring domain.

In the literature, many studies investigate new and improved techniques for various problems of AML. This chapter will focus on reviewing and analysing machine learning approaches for transaction monitoring in the literature. Section 2.2 provides the research methodology and Section 2.3 presents the theory and selection of the papers. Section 2.4 presents the results and discussion about the pre-processing methods utilised on the transactional data. Section 2.5 presents an examination of the datasets and feature selection processes. The machine learning algorithms for transactional monitoring are analysed in Section 2.6. The evaluation metrics are reviewed in Section 2.7 along with the categorisation of the papers according to their evaluation levels. Section 2.8 lists the shortcomings and the direction for future research. Finally, the summary and future work is presented.

2.2 Material and Methods

2.2.1 Search Strategy and Data Sources

This section explains how the research is conducted. A systematic literature review assists in identifying, selecting, and evaluating research published on a given topic, research area or research question. The systematic review was conducted in accordance with the approach outlined by Keele (2007). Important steps are shown in Figure 2.2, which include defining the

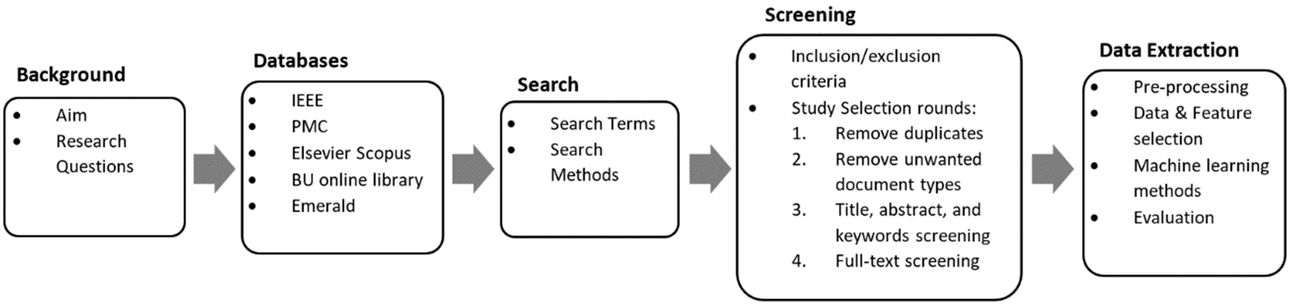


Figure 2.2: Systematic mapping process.

research questions, search strategy, inclusion and exclusion criteria, study selection, and data extraction method (Keele, 2007).

After preliminary searches about money laundering and machine learning in various electronic databases, the following data sources are selected: Institute of Electrical and Electronics Engineers (IEEE), PubMed Central (PMC), Elsevier Scopus, Bournemouth University (BU) online library and Emerald. The databases are selected as they cover the required relevant topics and are easy to access and acquire full-text articles. The Emerald database is included as it contains the Journal of Money Laundering Control which is directly related to the topic area. Springer was not selected as a standalone database since many of its relevant papers were already accessible through the BU online library, PMC, and Elsevier Scopus. Databases such as Google Scholar are not selected to perform searches as the resulting output is large and inaccurate which could have led to high probabilities of unrelated studies in the results.

The key terms are chosen in the attempt to include machine learning methods developed in literature aiming to enhance transaction monitoring. To decide the key terms the following guidelines by Keele, 2007 are utilised, 'derive key terms from research questions and study topics' and 'Identify synonyms, plurals, and related terms'. The selected key terms are money laundering, anti-money laundering, artificial intelligence, machine learning, transaction monitoring, detection, fraud, and crime.

2.2.2 Keywords

Advanced searches in the chosen databases are conducted by scanning the title, abstract, and keywords of studies for the predetermined key terms. Filters are applied to the searches in Elsevier Scopus, IEEE, and Emerald to remove unwanted document types, discussed further in the inclusion/exclusion section. PMC and the BU online library did not have the available filters to remove unwanted document types in the search, so they were removed in the study selection phase. The number of publications included for further review per database is shown in Table 2.1. The results of the search concluded a total of 366 papers. Combining the key terms with the "AND", "NOT", and "OR" logical operators meant the following is used for the advanced searches:

("Anti-money laundering" OR "money laundering") AND
("machine learning" OR "artificial intelligence") AND
("detection" OR "crime" OR "transaction monitoring" OR "fraud")
AND NOT ("Crypto*")

Table 2.1: Number of publications included for further review per database.

Database	Number of publications
Scopus	103
IEEE	71
BU online library	55
PMC	63
Emerald	74
Total	366

Table 2.2: The exclusion criteria for identified papers.

No	Exclusion Criteria
1	Transactional monitoring approaches that do not involve machine learning.
2	Approaches that do not use any transactional data.
3	Abstract-only papers as preceding papers, editorial, review papers, magazines, news, and books.
4	Articles without the full text available.
5	Duplicated and unrelated papers.
6	AML methods for Cryptocurrencies.

To target relevant studies and gather the necessary information to carry out this review, inclusion-exclusion criteria are set. The inclusion criteria are to select the studies that produce a machine learning method to detect money laundering in transaction monitoring. The exclusion criteria for this study are shown in Table 2.2. Methods produced for cryptocurrencies are excluded as they have a different data structure compared to traditional financial transactions.

2.3 Theory and Selection of the Papers

After the searches on the chosen electronic databases, all the studies are extracted to EndNote 20 to narrow down the results to relevant studies. Study selection is separated into four rounds. Round 1 is to remove all the duplicate papers. Unwanted document types stated in the exclusion criteria are removed in round 2. Round 3 is title, abstract, and keywords screening to remove studies that focus on AML methods for cryptocurrencies/blockchain and unrelated studies to the inclusion criteria. Full-text screening is the 4th and final round, the entire paper is downloaded, read, and examined concerning the inclusion and exclusion criteria. Studies that fall within the exclusion criteria or are unrelated to the inclusion criteria are removed.

A total of 366 studies were exported from digital databases, with 97 duplicates removed. In the second round, 21 studies were excluded, reducing the total to 255. During the third round, 93 studies progressed to the final full-text screening. At this stage, 10 papers on financial crime/fraud and 12 papers unrelated to money laundering were removed. Additionally, 15 papers that did not focus on transaction monitoring, 11 that did not propose a method, and 9 that introduced non-machine learning approaches were excluded. Ultimately, 36 studies, including manually added papers, were selected for the final analysis. A flowchart outlining the study selection process is presented in Figure 2.3.

To cover as many relevant studies as possible another step is taken. The following review

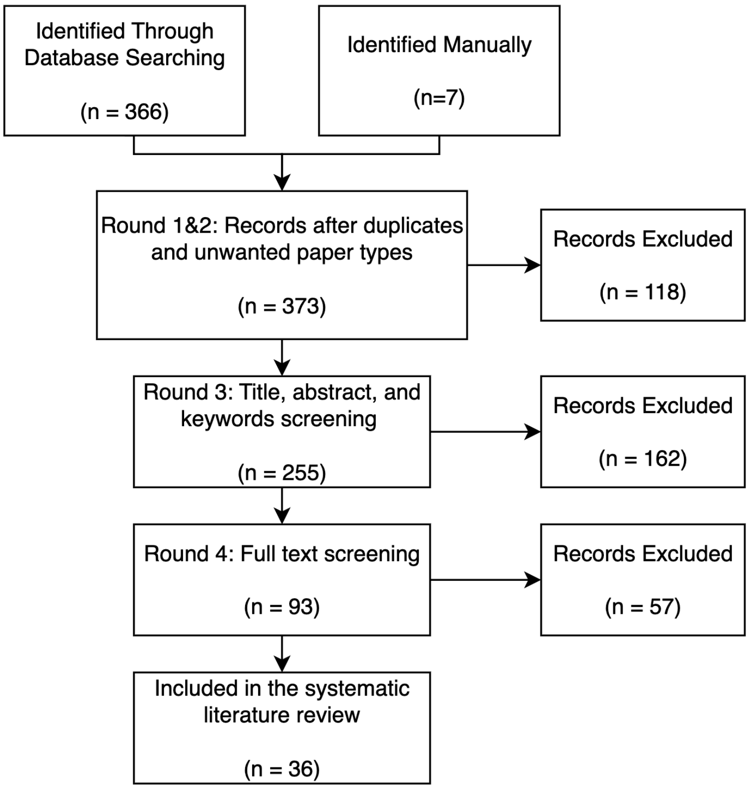

Identified Through Database Searching
(n = 366)
Identified Manually
(n=7)
Round 1&2: Records after duplicates and unwanted paper types
(n = 373)
Records Excluded
(n = 118)
Round 3: Title, abstract, and keywords screening
(n = 255)
Records Excluded
(n = 162)
Round 4: Full text screening
(n = 93)
Records Excluded
(n = 57)
Included in the systematic literature review
(n = 36)

Figure 2.3: Flowchart of the study selection process for the systematic literature review.

papers (Kute et al., 2021) (Z. Chen, Van Khoa, Teoh, et al., 2018), and (Leite, Albuquerque, and Pinheiro, 2019), found during the searches in the databases are used to manually identify studies to include in this literature review. This is done by scanning the reference list of the review papers. 7 studies are added to the final selection of papers, making 36 in total. The manually included papers did not appear in the searches due to not being published on the chosen databases or not including the ("machine learning" OR "artificial intelligence") key terms in the abstract, title, or keywords.

A data extraction form is designed to extract information from the selected studies and answer the research questions. The data extraction form is evaluated on a sample of included studies to assess if it works well. Standard information of studies like author's names, papers title, publication date, and journal/conference are included on the form. To answer the first research question, a column for pre-processing methods is included in the data extraction sheet. The next group of columns in the data sheet is the machine learning style (i.e., supervised, or unsupervised) and the method used. To gain a better understanding of feature selection techniques and data information the following columns are included in the data extraction form. Data type (i.e., real/synthetic/both), data source, data size (training/testing), suspicious transactions in data (training/testing), number of features used, and justification/explanation on feature selection. The final set of columns is evaluation methods, detection rate, and level of evaluation. The evaluation methods column includes all the techniques used within the paper. The detection rate is the true positive and false-positive rates of the method (TPR/FPR). The studie's level of evaluation is rated from 1-6 explained below:

- Level 1: No Evidence

- Level 2: Demonstration
- Level 3: Discussion via expert criticism or discussed advantages/disadvantages
- Level 4: Statistical results gained through experiments on synthetic data
- Level 5: Statistical results gained through experiments on real data
- Level 6: Further analysis (i.e., comparison of obtained results to an existing system in the industry)

Regarding the limitations of the study, this review study presents three primary limitations. Firstly, the search was restricted to four databases, which may not represent the entire existing literature. Secondly, the reliance on only the chosen keywords may have led to the exclusion of some relevant studies on machine learning methods for transaction monitoring. This may be due to the keywords missing in titles, abstracts, or keyword lists. The final limitation is about the language of the papers as only English articles were considered.

2.4 Results: Methods Used For Pre-Processing The Data

Data pre-processing is a vital stage in machine learning to clean, format, and organise raw data, enhancing its value and influence on model performance. In the context of this research, raw datasets refer to unprocessed transactional data held by the bank, typically consisting of basic transaction details such as sender, receiver, timestamp, location, amount, and transaction type. Processing this raw transactional data is crucial to attaining desirable machine learning results. The data must be structured precisely for effective interpretation by machine learning algorithms and cleaned to prevent errors, such as those caused by missing values. Additionally, customer behaviours can be analysed further to create new meaningful attributes, for instance, by assessing transaction frequency or periods between transactions. Processing transactional data can be challenging and time-consuming.

2.4.1 Data Structure Preparation

Data structuring preparation is the process of converting raw data into a particular format making it easier to apply and access. Transactional data for AML from financial institutions are stored in various formats as they are gathered from different databases (i.e., MySQL, Microsoft SQL Server, PostgreSQL, and Oracle). Transactional data is complex due to its high volume and number of features; therefore, data structuring is required to use the data for machine learning and to attain the best performance. Structuring the data is a difficult and time-consuming operation. For example, different types of transactions can be stored in separate tables such as cash, swift, cross border, domestic, and credit/debit card transactions. KYC information can also be stored in a different format and table. Combining two or more tables is complicated as there are millions of transactions and varying attributes. Matching transactions to the correct customer IDs and removing duplicates will be a lengthy task when combing tables. Out of the 36 studies analysed many conducted data structuring. Desrousseaux, Bernard, and Mariage (2021) gathered transactional data and external data from sources such as World-Check and sanction lists to combine the two and structure the information in a table format. The data was structured specifically to enable the application of machine learning methods.

2.4.2 Encoding

Encoding converts attributes from categorical to numerical so that machine learning techniques can interpret the data. Encoding techniques have to be selected with caution as they can impact the performance of machine learning methods (Pargent et al., 2022). A few studies in this review utilised encoding in their pre-processing phase. Zhiyuan et al. (2021) encoded the following three attributes, account type, product type, and business type using one-hot encoding and binary encoding. Both encoding methods resulted in the same outcome, therefore, binary encoding was chosen as it was computationally simpler. The methods of encoding were chosen as they were the most popular, but further research into other methods like the weight of evidence encoding could produce better results. Encoding is valuable in transactional monitoring as it allows KYC data to also be used to predict potential money laundering activities. Another encoding method used by Rouhollahi et al. (2021) is ".cat.code" which is a Python command. Transaction reference, transaction country, source, and destination countries are a few examples of the encoded variables.

2.4.3 Handling Missing Values

Some machine learning algorithms cannot process missing or poorly structured data, which can lead to errors in their predictions. Missing values can also negatively impact the performance of the machine learning method and cause problems when training the model. Transactional and KYC data collected from financial institutions can contain missing values for many reasons such as customers not sharing information. From the chosen studies, a few different approaches were taken to handle missing values. Zhiyuan et al. (2021) handled eight missing values by inputting the mean attribute value, whereas, the authors in Camino et al. (2017) and S. Raza and Sajjad Haider (2011) completely removed the transactions with missing values. Camino et al. (2017) filed the missing values for further investigation. Other ways to handle missing data can be investigated to improve the performance of the machine learning method such as predictive value imputation and iterative imputation (Saar-Tsechansky and Provost, 2007), (S. Zhang, Wu, and Zhu, 2010).

2.4.4 Standardisation/Normalisation

Standardisation and normalisation is a process to re-scale the data to a consistent range. It is beneficial when comparing attributes that have varying scales, as it prevents variables from over-influencing the results and generating a bias. In transactional data, many variables have different scales (i.e. transaction amount from \$1-\$500,000 (Rouhollahi et al., 2021)) so standardisation or normalisation can be used to improve the model's performance. Computation time and estimation error can also be decreased by standardisation and normalisation. Magomedov et al. (2018); Alshantti and Rasheed (2021) normalised their data whereas other studies ((Rouhollahi et al., 2021); Zhiyuan et al.) used standardising. In transactional monitoring, standardisation may be a better approach, as normalisation suppresses the effects of outliers (due to having a range between 0 and 1). Zhiyuan et al. (2021) stated that standardisation is a better approach than normalisation for fraud detection methods as it is crucial to maintain the original distance among the data points.

2.4.5 Feature Engineering

The procedure for feature engineering consists of combining, manipulating, and transforming raw data in various ways to engineer new features. The newly generated features aim to improve the performance of models by providing more desirable and meaningful data (Prado and Digiampietri, 2020). Feature engineering allows gaining the most out of the collected data. Transactional datasets contain many attributes that can be transformed to gain new information previously unavailable to enhance the accuracy of the model. Also, feature engineering can be used to simplify and reduce the number of included attributes, constructing more interpretable and usable attributes (Tundis, Nemalikanti, and Mühlhäuser, 2021). A majority of the studies collected in this systematic literature review conduct feature engineering, however, not much information is provided on the reasons why the attribute is created or used. Z. Gao (2009) conducted feature engineering to create the following attributes, withdrawal frequency, deposit frequency, and total frequency of transactions. These attributes help the author to identify two novel capital flows within a short time frame.

2.4.6 Data Consolidation

To gain better insight into the raw data, it can be split into various groups, making comparisons easier. Combining pieces of information such as time scope and account type can achieve consolidation. Ketenci et al., 2021 took a time series approach, splitting the data into quarterly and then daily windows (i.e., daily total incoming and outgoing funds). This splitting was intended to facilitate easier and more natural comparisons of transactions for suspicious clients. In contrast, another study (Kannan and Somasundaram, 2017) divided the data into individual, corporate, and financial institutions, focusing on transaction variables. Financial transaction data is large, making it computationally complex to make predictions. Methods such as Induced Ordered Weighted Average (IOWA) can reduce the size of a dataset without losing information. Hussain, Merigó, and M. R. Raza, 2022 combined weighted average (WA) and IOWA operators to produce an IOWAWA approach to reduce data dimensionality, which could be applied to transaction monitoring. Data consolidation can increase data quality, so further research on unexplored methods can improve predictions. However, this is a complex and time-consuming task, especially with transactional data.

2.5 Results: Data for Transaction Monitoring

The datasets and features used in transactional monitoring methods proposed in the literature are compared in Table 2.3. Real, synthetic, and mixed (a combination of real and synthetic) datasets have been used to conduct experiments, each influencing the performance, quality, and efficiency of machine learning methods in various ways. Real datasets have been used the most, although they can be difficult to acquire. The use of real datasets is beneficial as it closely represents real-world scenarios. However, evaluating results is challenging because real labelled data is limited and can be inaccurately categorised due to a lack of ground truths.

In addition, some real datasets may not include any money laundering transactions as they are very uncommon. For example, if a small real dataset is used there is a high probability no money laundering transactions will be included as the ratio of normal to suspicious transactions is extremely large (imbalanced data). To combat the label accuracy and imbalance problem mixed datasets have been applied during experiments by incorporating synthetic money laundering transactions. Full synthetic transactional datasets have also been adopted by authors to conduct tests as they can be accessed more easily and provide labels for all transactions. However, they may not provide a true representation of the method's ability in a real-world situation. Mixed and synthetic datasets can be time-consuming to create and may produce unrealistic real-world money laundering patterns.

Most of the experimental dataset sizes are below 1 million transactions, with over 50% containing fewer than 10,000 transactions. Only three of the studies specifying data size used over 1 million transactions to test their methods. The size of the testing dataset is crucial to assess the scalability of the proposed methods and to analyse their potential performance in a financial institution. A larger dataset is likely to yield more accurate results and provide a better demonstration of a method's real-world applicability. The time consumption of methods cannot be fairly compared as all the papers use different data sizes. The ratio between suspicious and normal transactions is important for conducting realistic tests. When financial institutions perform transaction monitoring, the number of money laundering transactions is very low compared to normal transactions. Many studies in Table 2.3 have a ratio above 10%, which is impractical as this would not be the case in a real situation.

There is a need for an easily accessible, high-quality base dataset for authors to test their methods. Currently, all methods are tested on different datasets with varying sizes and features, making comparisons impossible. A base dataset could also help identify problems with existing methods, allowing for further enhancement. At present, authors mainly develop new methods instead of improving existing ones due to a lack of access to datasets. Most datasets used for experiments are inaccessible, making the proposed methods difficult or impossible to reproduce. Additionally, some papers lack information on the dataset, data pre-processing, or feature selection process, further complicating reproduction and decreasing reliability. Financial institutions may be reluctant to use the proposed methods due to this lack of information and trust, despite some methods showing promising results. Testing methods on more than one dataset could prove their consistency and reliability, providing insight into how the method will react to different datasets and a better understanding of its potential deployment in the industry.

2.5.1 Feature Selection

Feature selection is a key process when building a machine learning model as it can significantly influence the accuracy of the method and reduce computational costs (Cai et al., 2018). It achieves these advantages by selecting the most appropriate and consistent features while removing the irrelevant ones. Most papers' feature selection methods included in this review have not been analysed thoroughly and explanations of the chosen features are lacking. Only

Table 2.3: Overview of the datasets used for transaction monitoring experiments(T/T denotes Training/Testing in feature names).

Paper	Data Source for Transactional Data	Data Type (T/T)	Data Size (T/T)	Suspicious Trans. (T/T)	% of Sus to Norm Trans.
Zhiyuan et al., 2021	Malaysian Financial Inst. & Univ. of Nottingham	Mixed/Mixed	(4011/1003)	(0/265)	26.42%
Rouhollahi et al., 2021	Unnamed Financial Institution	Real/Real	(-/-)	(-/-)	-
Alshantti and Rasheed, 2021	DNB Bank	Real/Real	(9128/2282)	(913/228)	10%
Desrousseaux, Bernard, and Mariage, 2021	-	Real/Real	(-/710,000)	(-/-)	-
J. D. J. Rocha-Salazar, M. J. Segovia-Vargas, and M. D. M. Camacho-Miñano, 2021	Financial Inst. in Mexico	Real/Real	(30,278/1600)	(-/370)	23.13%
Stojanović et al., 2021	www.kaggle.com	Real/Real	(199,365/85,442)	(0/489)	0.57%
Stojanović et al., 2021	www.kaggle.com	Synthetic/Synthetic	(4,453,834/1,908,786)	(0/8207)	0.43%
Stojanović et al., 2021	www.kaggle.com	Synthetic/Synthetic	(416,250/178,392)	(0/7195)	4.03%
Tundis, Nemalikanti, and Mühlhäuser, 2021	www.kaggle.com	Synthetic/Synthetic	(3,600,000/2,400,000)	(-/-)	-
Ketenci et al., 2021	AkBank	Real/Real	(6680/4263)	(1787/995)	23.43%
Guevara, Garcia-Bedoya, and O. Granados, 2020	Non-bank correspondent in Colombia	Real/Real	(-/-)	(-/-)	-
Jullum et al., 2020	DNB Bank	Real/Real	(13,782/4967)	(0/147)	2.96%
Amr Ehab Muhammed Shokry, Mohammed Abo Rizka, and Nevine Makram Labib, 2020	AML Sim by (Weber et al, 2018)	Synthetic/Synthetic	(-/-)	(-/-)	-
Y. Zhang and Trubey, 2019	US Financial Institution	Real/Real	(3922/2157)	(28/6)	0.28%
Tai and Kan, 2019	Bank SinoPac	Real/Real	(739,142/-)	(-/-)	-
D. Huang et al., 2018	ICIJ Offshore Leaks Database from IKnow.com	Mixed/Mixed	(0/573,113)	(-/-)	-
Magomedov et al., 2018	-	-	(993/437)	(516/199)	45.54%
Camino et al., 2017	Private Transaction Monitoring Company	-	(-/-)	(-/-)	-
Michalak and Korczak, 2011	Produced by author	Synthetic/Synthetic	(2,854,965/2,625,671)	(3890/3982)	0.15%
X. Liu and P. Zhang, 2010	Financial Institution in Shanghai	Real/Real	(-/122,783)	(-/1797)	1.46%
Z. Gao, 2009	Unnamed Financial Institution	Mixed/Mixed	(34,343/Na)	(-/-)	-
Jun and Jian, 2005	Agriculture Bank in China	Mixed/Mixed	(318,070/-)	(30/-)	-
Usman, Naveed, and Munawar, 2023	www.kaggle.com	Synthetic/Synthetic	(-/-)	(-/-)	-
Phyu and Uttama, 2023	www.kaggle.com	Synthetic/Synthetic	(-/-)	(-/-)	-
Ruchay et al., 2023	www.kaggle.com	Synthetic/Synthetic	(198,608/85,118)	(344/148)	-
Hayble-Gomes, 2023	Created by the Author	Synthetic/Synthetic	(90,963/60,642)	(-/-)	22%
Karim et al., 2024	www.kaggle.com	Synthetic/Synthetic	(-/-)	(-/-)	2%
Koo, Park, and Yoon, 2024	South Korean Bank	Real/Real	(60,000/6000)	(-/-)	-
Xia et al., 2024	www.kaggle.com	Synthetic/Synthetic	(-/-)	(-/-)	-
Z. Chen, Khoa, et al., 2014	Bank in Malaysia	Real/Real	(2035/665)	(31/14)	2.11%
Keyan and Tingting, 2011	Agriculture Bank in China	Mixed/Mixed	(32,000/-)	(1930/-)	-
Khac and Kechadi, 2010	CE Bank	Real/Real	(3000/-)	(8/-)	-
Wang and Dong, 2009	Unnamed Financial Institution	Real/Mixed	(-/65,001)	(-/60)	0.09%
Lv, Ji, and J. L. Zhang, 2008	Unnamed Financial Institution	Real/Mixed	(270/90)	(70/40)	66.67%

31.1% of the studies included in this review paper conducted and discussed a feature selection process.

The most popular approach to determining the attributes was the filter method. The filter method identifies the best features based on statistics (Cherrington et al., 2019). Wrapper methods and embedded approaches have also been used. Wrapper methods see the feature selection process as a search problem and use trial and error to discover the best subset of the features (G. Chen and J. Chen, 2015). The feature selection process is incorporated into the machine learning technique in embedded approaches (Chandrashekar and Sahin, 2014).

The Weight of Evidence algorithm is used by Stojanović et al., 2021 to calculate an information value for each attribute and rank it to decide which variables will be included. Attributes are placed in one of the four variable predictiveness categories created. The four categories and information value boundaries set are developed for credit scoring but are used for fraud detection by the author. In (Alshantti and Rasheed, 2021), the authors used a combination of feature selection algorithms to identify the most influential features. A mean ranking is taken from the five algorithms to select the features. The algorithms used are L2 Regularisation, Gini impurity, ANOVA F-Score, Fisher Score, and Fast Correlation-Based Filter Solution (FCBF). Feature selection was conducted by applying PCA by Larik and S. Haider, 2011. The results showed that 80% of the total variance in the dataset was covered by nine out of sixteen features. Guevara, Garcia-Bedoya, and O. Granados, 2020, and Y. Zhang and Trubey, 2019 also used a filter method, however, did not use algorithms and analysed the data instead. The features were chosen based on a thorough data examination. An embedded method was used by Zhiyuan et al., 2021 to detect money laundering transactions by utilising autoencoders. Autoencoders integrate the feature extraction process as a part of the learning algorithm as it learns the most important features of normal transactions to identify suspicious ones. The author in Tai and Kan, 2019 used a wrapper method called top-down feature elimination and decided to include 18 features. The study lacks information on how the top-down feature elimination method was performed. Also, this approach can be very time-consuming and inefficient.

2.6 Results: Machine Learning Methods

Many different machine learning approaches have been proposed for transaction monitoring. Figure 2.4 illustrates all the methods discussed in the literature reviewed in this paper. The types of machine learning techniques used by the authors included in this study are presented in Figure 2.5. The most popular methods are Bayesian Algorithms, SVMs, XGBoost, and Random Forests.

2.6.1 Neural Networks

Artificial Neural Networks (NNs) are a deep learning approach in machine learning. An ANN is a computational model consisting of many simple, connected processors called neurons, each producing a sequence of real-valued activations (Schmidhuber, 2015). NNs are inspired by the way a human brain works, sending signals from one neuron to another. The structure includes an input layer, hidden layers, and an output layer, all linked to create an NN (Schmidhuber, 2015). Selected attributes of the dataset are inputted into the first layer and then sent to the

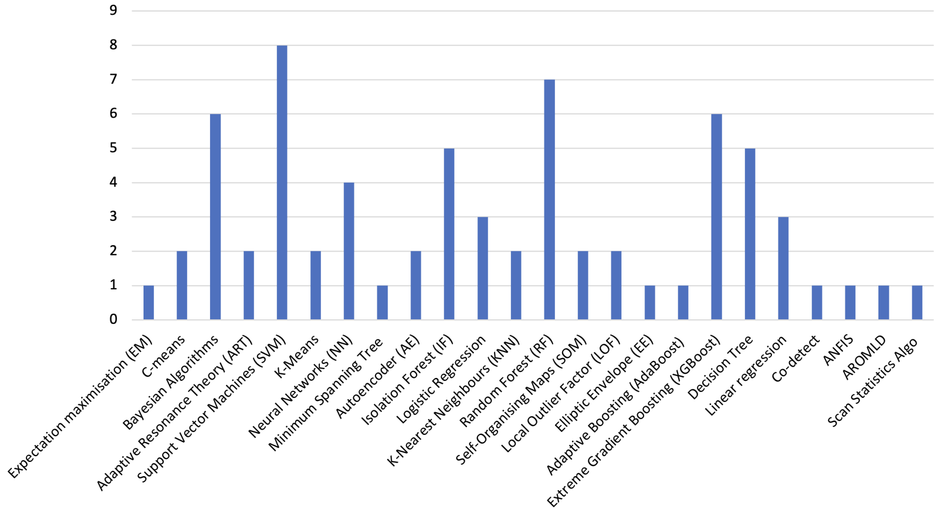


Figure 2.4: Machine learning techniques used for transaction monitoring.

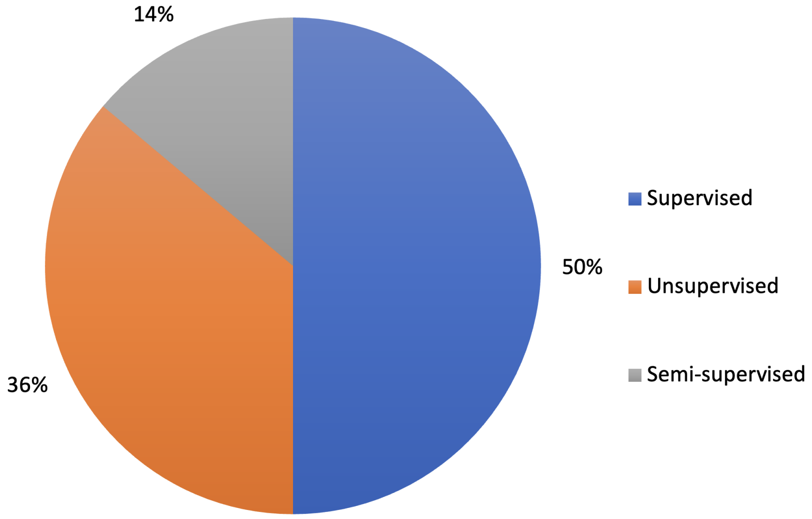


Figure 2.5: Number of published studies based on the machine learning approach used.

next layer (hidden layers). Embedding of categorical variables is required to input the data into the first layer. Neurons are activated based on set thresholds and weighted connections, which then go on to activate other neurons. Finally, the output layer returns a prediction value for a data point. NNs can be suitable for transactional monitoring due to their ability to solve problems and recognise patterns in complex and non-linear datasets. They can also process unclear and incomplete data, which may be beneficial to financial institutions that have incomplete data. However, financial institutions may find it difficult to adopt an NN-based method as they are challenging to interpret. The output of an NN relies heavily on data quality; therefore, concrete money laundering data is required for transactional monitoring, which is difficult to acquire.

Lv, Ji, and J. L. Zhang, 2008 built a transaction monitoring method using a type of NN called a radial basis function (RBF). To calculate the parameters of the hidden layer in the RBF model, the APC-III clustering was used, which sped up the learning duration. The weights of links between the hidden and output layers were updated by the recursive least squares (RLS) algorithm. The author decided to utilise the RLS algorithm to increase the convergence speed, as the gradient descent algorithm performed poorly during tests. The RBF method was compared and proven to outperform the SVM and an outlier detection method regarding true and false positive rates. The method was tested on a small dataset and used an unrealistic ratio of money laundering transactions to normal ones. Forty out of ninety transactions included in the testing dataset were money laundering transactions. The false positive rate was low, however, this is due to the data size and ratio, as over half of the transactions in the data are classified as suspicious. Also, only the following variables were used, transaction amount, frequency of deposits, and frequency of withdrawals. More variables can be considered for future research to increase the accuracy and reliability of the method.

Y. Zhang and Trubey, 2019 conducted experiments using multiple algorithms to detect potential money laundering transactions. A single hidden layer feedforward network (SLFN) was proven to be the best performing after conducting experiments on the following methods: Bayes logistic regression, decision tree, random forest, and support vector machine. The author found that NNs can overfit to the training data when the focus is solely on reducing the error to near-perfection. The methods were evaluated by analysing their ROC graphs, under- and over-sampling methods, graphic illustration, and regression analysis. The results of the SLFN method heavily rely on the quality of the labelled money laundering data collected by the author.

Jamshidi et al., 2019 proposed a method based on the Adaptive Neuro-Fuzzy Interface System (ANFIS) to detect suspicious customers by analysing transactions. The method included three inputs and then produced a result. The inputs were "Number of Exchange", "Standard Exchange", and "Real Exchange". The study lacked information on the process of the method. Also, the experiment was done on a small dataset with unclear and minimal evaluation.

Autoencoder

Unsupervised neural network techniques called Autoencoder and Variational Autoencoder (Baldi, 2012) are proposed to enhance AML by Zhiyuan et al., 2021. In this study, only normal transactions are used to train the models. Therefore, the methods only learn to recreate normal transactions. During the testing stage, normal and suspicious transactions are utilised and the recreation error for each transaction is calculated. If the error is above the set threshold the transaction is labelled as suspicious. The Recall-First Threshold (RFT) is used to optimise the threshold value to ensure the method is working efficiently and to enhance the detection rates. The experiments were conducted on a real and mixed dataset. The mixed dataset included synthetic money laundering transactions generated by the Wasserstein Generative Adversarial Network (WGAN). The goal of using WGAN was to produce a more balanced dataset. The mixed dataset reduced the false positive rate drastically from 19% to 8%. The Autoencoder using the mixed dataset outperformed the rest of the experiments. However, even after applying cross-validation, the mixed dataset was prone to overfitting. In addition, although the method reduced the false positive rate severely, the precision was too low. Further research on several other autoencoder methods can be conducted to gain more desirable results (i.e. variational ladder autoencoder). For example, a memory-augmented autoencoder is used by authors in H. Gao et al., 2022 to detect anomalies in time series data which can be applied to monitor transactions.

Koo, Park, and Yoon, 2024 proposed a model that incorporates autoencoders within a risk-based framework to detect suspicious activities. This model is trained on a dataset comprising approximately 60,000 transactions from a major bank in South Korea. The model utilises an unsupervised learning approach to identify complex and nonlinear dependencies that indicate potential financial misconduct. The proposed method is assessed alongside Random Forest, with a detailed comparison of the advantages and disadvantages of each. Moreover, the model surpassed Random Forest in terms of accuracy. By applying a risk-based approach, the model efficiently detects and prioritises suspicious transactions, reducing the operational burden on financial institutions by minimising the need for manual reviews. The study emphasises the model's adaptability and potential for real-time application, highlighting its capability to dynamically reflect current transaction patterns and behaviours. This adaptability makes it a valuable tool for enhancing the scalability and accuracy of fraud detection systems.

2.6.2 Support Vector Machines

In Cortes and Vapnik, 1995 paper, the authors introduced a supervised machine learning technique to solve regression, classification, and outlier detection problems called SVM. Data points are separated by a hyperplane and the SVM algorithm assists in finding the best possible one in N-dimensional space (N represents the number of features used). The goal is to find the optimal hyperplane which is done by choosing the plane with the maximum distance to the nearest data points in each cluster. After the optimal plane is found using the training data, the testing data is classified depending on the side of the plane it lies within. SVM is effective when dealing with high-dimensional datasets which will be advantageous for transactional monitoring. However, SVMs under perform on large datasets and cannot handle noise very well, both necessary for transactional monitoring. Also, money laundering and normal transactions overlap most of the time which will make it difficult for SVMs to create a boundary splitting the data.

Keyan and Tingting, 2011 proposed an SVM method developed based on the study by Jun and Jian, 2005. Cross-validation was incorporated into the method to identify the best parameters. The objective of using cross-validation was to improve the results of the previous experiment and produce a more efficient method. In this study, the K-CV method was used. The K-CV method measures the training set's accuracy for multiple parameters (c and g), and the optimal parameters were chosen based on the highest classification accuracy. Two experiments were conducted. The first aimed to understand customer behaviour using the SVM method, setting parameters by trial and error. The second focused on optimising the parameters using the cross-validation technique. The results of the experiments showed an improvement in the detection rate and accuracy of the SVM method when incorporating cross-validation.

2.6.3 Clustering Approaches

The authors used K-means in (Khac and Kechadi, 2010) to cluster transactions and label them as normal or suspicious. A neural network was trained using the labelled clusters. As the data used by the author was imbalanced and did not have enough fraudulent transactions, more were created, based on the existing ones. The proposed method was evaluated on new transactions and outperformed the bank's existing approach in terms of running time. The proposed method required expert knowledge to label the various clusters, which could be costly and time-consuming for financial institutions. Also, as stated by the author, an approach to decide on the number of clusters to create to optimise results could be studied further.

In addressing financial fraud, Xia et al., 2024 proposed the ENKMRH model, a selective ensemble prediction method combining K-means++ and the Refractive Inverse Learning Harris Hawks Optimisation (RILHHO) algorithm. This innovative approach ensures diversity and efficiency in base learner selection, significantly improving the prediction performance and generalisation capabilities of the ensemble system. Applied to three different fraud datasets, the ENKMRH model achieved an accuracy rate of 93.80% on the IBM AML dataset. The study highlights the importance of integrating advanced algorithms with risk management strategies to tackle modern financial fraud challenges. Future research aims to refine feature engineering methods and address data privacy concerns, further advancing the effectiveness of fraud risk prediction.

J. D. J. Rocha-Salazar, M. J. Segovia-Vargas, and M. D. M. Camacho-Miñano, 2021 proposed a three-phase method. First risk ratings were given to attributes in the dataset using fuzzy logic. Then the C-means algorithm was used to create risk clusters. In the last phase, an anomaly indicator identified transactions by measuring the difference between the variables. Further research could be done to avoid the following issues. The detection rate of the proposed method is exceptional, however, the test is conducted on a small dataset (1600 transactions). The detection rate may not be an accurate representation of the method in real-world use. Furthermore, the ratio of money laundering transactions to normal transactions included is unrealistic (20% of the testing data set is money laundering transactions).

In an attempt to detect potential money laundering transactions, a study proposed a Minimum Spanning Tree (MST) Clustering algorithm (Wang and Dong, 2009). By separating the dataset into upper, middle, and lower datasets, the author enhanced the MST algorithm. The method was tested on both real and synthetic datasets, with 60 synthetic money laundering transactions included. The detection of suspicious transactions was proven to be exceptional; however, synthetic money laundering data may not represent real-world data very well. Although the detection of anomalies was good, no information on false positives was included. To improve the MST clustering algorithm further, the value for k (the number of clusters) could be optimised instead of using a trial-and-error approach.

To detect outliers in transactions the authors in S. Raza and Sajjad Haider, 2011 proposed a method that uses fussy c-means and dynamic Bayesian networks called SARDBN (Suspicious Activity Reporting using Dynamic Bayesian Network). The method has three major steps, clustering, learning using DBN, and anomaly detection. A formula called AIRE (Anomaly Index using Rank and Entropy) is created to rank and group the data points. It is difficult to evaluate the models' performance and find the optimal quantity of clusters to generate.

2.6.4 Graph Analysis

Usman, Naveed, and Munawar, 2023 explored the application of a graph-based machine learning model using the synthetic AMLSim dataset. The study employed a graph convolutional network (GCN) for semi-supervised learning and classification of financial transactions, showing that the GCN significantly outperforms traditional rule-based systems. However, it used accuracy as the key evaluation metric, which is not ideal for transaction monitoring due to the imbalanced nature of the dataset. Future work could apply this approach to real-world datasets to better evaluate its performance in practical scenarios.

Another study (Michalak and Korczak, 2011) proposed a machine learning method to mine transaction graphs. A graph structure learning method is constructed, which can select potential money laundering transactions. The method is trained on transaction graphs that have been previously annotated. Finally, the method is used on an unannotated graph, and potential money laundering transactions are flagged for further investigation. The results of the method were promising, however, improvements in precision could be made, as mentioned by the author. Additionally, evaluating the method on real data is another crucial area for future research.

Another graph analysis approach was suggested in (D. Huang et al., 2018), called CoDetect. Real-world and synthetic data were combined to evaluate the CoDetect method's efficiency. The detection accuracy of CoDetect is determined by evaluating the accuracy of the similarity matrix and feature matrix. This measure may not be a very accurate representation of the method in real-world use. The CoDetect methods approach is easily interpretable, hence, making it easier for financial institutions to adopt.

In the study by Karim et al., 2024 an approach utilising semi-supervised graph learning techniques on large-scale financial transaction graphs is proposed. This research utilises four datasets, including AMLSim, Elliptic, IBM AML, and SynthAML, to evaluate the scalability and practical applicability of the models. The methods employed, integrate both pipeline and end-to-end settings for graph learning. In the pipeline settings, graph embedding models are initially trained to generate node embeddings, which are then used alongside topological graph features to train classifiers. In the end-to-end setting, models like SkipGCN, FastGCN, and EvolveGCN perform node classification directly without the need for separate classifiers. The GCN models slightly outperform XGBoost for two datasets but perform worse for the other two. Additionally, the study explores the interpretability and transparency of AML models by offering both local and global explanations for predictions and discusses how these models can be integrated into real-world financial systems.

2.6.5 Self-Organising Map and Adaptive Resonance Theory

Self-organising map (SOM) and Adaptive Resonance Theory (ART) are both unsupervised neural network-based methods (Rui and Wunsch, 2005). ART clusters the data and once new data points are inputted into the algorithm, they either join an existing cluster (if attributes are similar) or start a new cluster. SOM converts a high-dimensional dataset, by clustering and mapping, into a lower dimension while sustaining the dataset's core principles to simplify the problem's interpretability. As transactional monitoring is highly complex SOM could be a good approach. ART could be beneficial in AML as similar transactions will be clustered together which can be investigated further to identify suspicious transactions. Also, new money laundering approaches can be identified if they do not fit into existing money laundering clusters.

SOM was proposed in (Alshantti and Rasheed, 2021) with an extra step of clustering to enhance transaction monitoring. SOM produces an inter-neural distance of all neurons that range between 0 and 1. Then depending on the distance, the neurons are placed into one of the five risk-level clusters. The experiment was conducted on a real and labelled dataset, however, the number of transactions included is small compared to real-world applications. The proposed method reduced the false positive rate drastically (6.2%) compared to existing rule-based methods, however, the true positive rate (65.5%) suffered. This has a far worse impact on financial institutions as many potential money laundering transactions will be classed as normal. The author states that the efficiency of the SOM method reduces with imbalanced data which is a problem for transactional monitoring as the proportion of money laundering transactions is far less than normal transactions. The reliability of the method relies heavily on how well the data was labelled by the bank (DNB Bank). Further research could be done on a larger data set that considers more attributes.

Desrousseaux, Bernard, and Mariage, 2021 went a step further and proposed a method that utilises SOM and ART. A variation of ART called Fuzzy ART (fzART) was used. First SOM calculates the weights for each data point and displays it on a U-matrix. The weights are inputted into fzART where clusters are produced with categorical labels. Combining SOM and fzART reduces the execution time of the method considerably compared to executing each method separately. Nine clusters are created and analysed; however, it is difficult to tell the accuracy of the model as the data is unlabelled. Over 21% of the transactions are in an unclear cluster which will be a problem for financial institutions when deciding to report a transaction. Furthermore, the other clusters do not give a clear indication of whether a certain transaction should be alerted or not, therefore, an in-depth analysis of each transaction will still need to be done by financial institutions. This will lead to high costs which is already a major problem in AML. Future research could focus on adding an extra procedure to solve this problem.

2.6.6 Isolation Forest

The authors in (F. T. Liu, Ting, and Zhou, 2008) developed an unsupervised machine learning outlier detection algorithm called Isolation Forest, based on decision trees. Unlike most algorithms, which separate outliers after constructing an average of the data, Isolation Forest immediately isolates unusual data. Anomalies are identified by an ensemble of tree structures, starting from a randomly chosen data point and decided on randomly chosen attributes. Data points closest to the origin of the tree are given a larger anomaly score. Isolation Forest identifies anomalies based on the belief that they have different attribute values and cover a small proportion of the dataset. Isolation Forest can be applied to large datasets and achieve an exceptional detection rate on anomalies within a good time frame, which is necessary for transactional monitoring. However, Isolation Forest's false positive rate could be negatively affected, as many normal transactions are similar to suspicious transactions that are well hidden among them.

The Isolation Forest and One-Class SVM techniques were compared in (Amr Ehab Muhammed Shokry, Mohammed Abo Rizka, and Nevine Makram Labib, 2020) to detect suspicious activities by investigating the amount and time of transactions in a synthetic dataset. Isolation Forest was deemed the better method as it had a shorter computation time, used less memory, and achieved better accuracy. However, not much information or explanation about the accuracy of the model is provided, other than stating that specialists confirmed Isolation Forest had better accuracy. Another study (Guevara, Garcia-Bedoya, and O. Granados, 2020) also experimented with the Isolation Forest technique but investigated the withdrawal information of transactions instead. Guevara, Garcia-Bedoya, and O. Granados, 2020 used a real dataset that included more information, due to more in-depth variables, compared to (Amr Ehab Muhammed Shokry, Mohammed Abo Rizka, and Nevine Makram Labib, 2020) (9 features were used compared to 4). Hence, the results in Guevara, Garcia-Bedoya, and O. Granados, 2020 may be a better representation of the method in a real-world application. Stojanović et al., 2021 compared the Isolation Forest method to the Local Outlier Factor and Elliptic Envelope methods using three datasets. Isolation Forest performed the best regarding sensitivity, specificity, and ROC in two of the datasets, while Local Outlier Factor performed the best in the other.

2.6.7 Local Outlier Factor

A method to find anomalies in datasets called the Local Outlier Factor (LOF), was introduced by (Breunig et al., 2000). LOF is an unsupervised machine learning algorithm that calculates a local outlier score concerning a specific instance and its k-nearest neighbours. A data point is labelled as an outlier if it has a considerably lower density compared to its neighbours. The parameter k must be set cautiously, as a small value for k will lead to a high false-positive rate as anomalies will be detected in many small areas. A large value for k could lead to a low true positive rate as a lot of suspicious transactions will be missed and only global anomalies will be detected. LOF algorithm, with a good k input, is efficient at finding local anomalies which can be beneficial in transactional monitoring as many money laundering transactions are submerged in between multiple normal transactions. At financial institutions millions of transactions get monitored to prevent money laundering, therefore, scalability could be a problem if LOF is adopted as it requires a lot of computational power and memory to be used.

Z. Gao, 2009, produced a cluster-based local outlier detection factor (CBLOF) algorithm to identify suspicious money laundering transactions within a real dataset that included synthetic suspicious transactions. 33 of the 40 synthetic money laundering transactions are identified by the method. However, the study lacked evaluation making it difficult to assess the effectiveness of the detection rate (i.e., no information about the false-positive rate). In addition, the detected suspicious transactions are created and not real so the results may not be a true representation of real data. The method in (Z. Gao, 2009) only included a few behavioural attributes of the transactions (withdrawal and deposit frequency) and left a future research area to be explored using more attributes. Stojanović et al., 2021 conducted an experiment using the LOF algorithm on a synthetic dataset that outperformed the Isolation Forest algorithm. The evaluation level of the study Khac and Kechadi, 2010 was higher than Z. Gao, 2009 study as more statistical evaluation methods were used (i.e. ROC). Also, more behavioural features were utilised compared to Z. Gao, 2009, but the dataset lacked a lot of information. Therefore, further research could be done using the LOF algorithm on a detailed dataset.

2.6.8 One-Class Support Vector Machines

One-Class Support Vector Machine is an unsupervised machine learning technique (Chang and Lin, 2011). The One-Class SVM computes its boundaries by analysing normal instances and does not need labelled data, unlike the traditional SVM method. Instances beyond the boundaries are regarded as anomalies. For transactional monitoring, the One-Class SVM characteristic of not needing labelled data is beneficial as accurately labelled data is difficult to acquire in AML. Also, the One-Class SVM method reduces the number of anomalies to detect due to its approach which could reduce the false positive rate in transactional monitoring, however, this may come at the cost of reducing the true positive rate.

The One-Class SVM algorithm is used on heterogeneous datasets, made possible by creating an RBF kernel function based on HVDM distance (Jun and Jian, 2005). The author selected the SVM algorithm due to its strong performance in handling high-dimensional transactional datasets, particularly its ability to effectively separate classes in complex feature spaces using kernel functions. The false-positive rate (5.4%) of the method is a huge improvement on existing methods, but the accuracy is low and could be enhanced (69.13%). The two parameters in the approach (parameters C - incorrectly classification punishment parameter and g - control factor) were adjusted using a trial-and-error technique. To improve results a method to optimise the parameters could be done.

Additionally, SVM's robustness to overfitting, especially in cases where the number of features exceeds the number of samples, makes it a suitable choice for financial transaction analysis.

Guevara, Garcia-Bedoya, and O. Granados, 2020 and Amr Ehab Muhammed Shokry, Mohammed Abo Rizka, and Nevine Makram Labib, 2020, both used One-Class SVM to detect suspicious anomalies within transactional data. Guevara, Garcia-Bedoya, and O. Granados, 2020 used a real dataset concerned with withdrawal information and the authors in Amr Ehab Muhammed Shokry, Mohammed Abo Rizka, and Nevine Makram Labib, 2020 utilised a synthetic dataset analysing the amount and time attributes of transactions. In (Amr Ehab Muhammed Shokry, Mohammed Abo Rizka, and Nevine Makram Labib, 2020) experiments showed that the One-Class SVM method required too much memory and training/prediction time. Furthermore, the method's ability to detect suspicious transactions was below requirements.

2.6.9 Decision Tree

The authors in (Phyu and Uttama, 2023) analysed the classification performance on the synthetic IBM AML dataset using several machine learning models, including k-nearest neighbour (KNN), decision tree (DT), support vector machine (SVM), and naïve bayes (NB). Their results indicated that supervised machine learning models, particularly the decision tree, with an accuracy of 87.6%, were effective in classifying money laundering transactions, outperforming KNN (87.2%), SVM (84.4%), and NB (71.3%). The study was further extended by examining typological interactions and classifying them individually as features, with the mutual typology consistently yielding the highest accuracy among single typologies. However, the use of accuracy as the evaluation metric is not ideal for imbalanced datasets like those in transaction monitoring. Furthermore, the study could have benefited from including other advanced supervised methods such as XGBoost or Random Forest for a more comprehensive analysis.

In another study, Hayble-Gomes, 2023 explored the use of Logistic regression and Decision Trees to identify relevant attributes for filing SARs by analysing historical customer transaction data in retail banking within the USA. Utilising machine learning classifiers, the study focused on metrics such as accuracy, TPR, FPR, and FNR. The decision tree model was insightful, highlighting key variables like prior SAR reports, customer age, duration with the bank, transaction amount, and alert score. Although the research demonstrated that machine learning could enhance transaction monitoring, it concluded that identifying SAR features alone is insufficient for comprehensive monitoring. The study underscores the need for continuous updates to AML scenarios and compliance programs, advocating for the integration of predictive modelling with traditional rule-based systems to optimise transaction monitoring without compromising compliance integrity.

2.6.10 Random Forest

Random Forest is a supervised machine learning algorithm that can be used for regression and classification problems, first introduced by Breiman, 2001. The method's final output is decided by a combination of multiple decision tree results. The type of problem affects how the final output is decided, for classification problems a majority vote is used, and for regression, an average is calculated. When creating each decision tree bagging is utilised to enhance the predictions of the method by injecting randomness. Bagging assists in diversifying decision tree outcomes by allowing the tree to input randomly selected data from the dataset. Due to this, random forest reduces the risk of overfitting. The Random Forest method can have long consumption times when dealing with large datasets which could be a problem for transactional monitoring.

Ketenci et al., 2021 used Random Forest to develop a transaction monitoring system focusing on the time-frequency attributes of customers. The method is trained on previously labelled money laundering transactions. A threshold is set and transactions with values above the threshold are deemed suspicious. The author suggests that the time-frequency characteristics of customers will allow the system to detect suspicious transactions. A time-frequency domain representation is created, where daily transactions are analysed within a quarterly sliding time frame. By including the time-frequency attributes the study showed that it simplified the attribute selection process and improved the performance. The dataset used was small, and more experiments could be done using a larger dataset to get a better idea of the performance of the method. Also, the ratio of suspicious transactions to normal transactions is unrealistic.

Tundis, Nemalikanti, and Mühlhäuser, 2021 experimented on and compared the following supervised machine learning algorithms, Decision Tree, SVM, Linear Regression, Naïve Bayes, and Random Forest. Random Forest was the best performing algorithm in classifying money laundering transactions. A synthetic dataset was used to conduct the experiments with accuracy, recall, precision, and F1-score evaluation metrics used to compare the methods. Further analysis was done to identify the most influential and important attributes. The author identified balance_difference_90, balance_difference_60, and outgoing_foreign_amount_30 as the top 3 most important attributes. Although the feature analysis can help with interpretability, financial institutions may still be hesitant to use the Random Forest method, due to alternative approaches. Another study (Ruchay et al., 2023), also experimented with the Random Forest algorithm on an imbalanced dataset known as CreditCardFraud. The researchers used a combination of the Tree-based Pipeline Optimisation Tool (TPOT) and Random Forest to create their model. This approach achieved a detection accuracy of 99.99% for fraudulent transactions, which was confirmed through 10-fold cross-validation. While this model reached the highest accuracy, it was slightly surpassed by other models in terms of precision and recall. Additionally, the incorporation of the Tomek links resampling algorithm greatly improved the reliability of detecting fraud in scenarios where the data classes are imbalanced.

2.6.11 Boosting Algorithms

Boosting algorithms are an ensemble approach that is implemented alongside a weak algorithm to enhance it and attain better results. Extreme Gradient Boosting (XGBoost) and Adaptive Boosting (AdaBoost) have both been used for transaction monitoring. XGBoost utilises a gradient boosting structure and improves upon it. The improvements are done through parallel processing, handling missing data, regularisation, etc. AdaBoost reallocates weights for every case of the weak learning algorithm. Incorrectly classified cases are given a higher weight, whereas cases that are classified accurately are given a lower weight.

Jullum et al., 2020, proposed an XGBoost method to detect suspicious transactions and compared it to a bank's existing method. The method outperforms the bank's approach in the following metrics, ROC-AUC, Brier score, and PPP (metric created by the author). This study included transactional alerts that did not lead to an AML report when testing and creating the method to enhance results, which has not been done in any other study. Stojanović et al., 2021 conducted three experiments on three different datasets (two synthetics, one real) comparing the results of Random Forest, AdaBoost, and XGBoost. XGBoost gave the best results in two of the experiments and AdaBoost in the other. Sensitivity (true positive rate) and specificity (true negative rate) were used to decide on the best-performing methods. The false positive rate is not calculated or used to compare the methods which would be a good metric to evaluate the efficiency of the methods. As false positives are an issue for financial institutions more research could be done on the proposed methods in this area.

2.6.12 Other Methods

An Expectation Maximisation (EM) approach for AML is conducted in (Z. Chen, Khoa, et al., 2014). A real dataset from a Malaysian bank is used to perform the experiment. The data was consolidated into the following periods, daily, weekly, and monthly. The daily and weekly time frames were where most of the suspicious transitions were detected as shown during the experiment and stated by the author. The EM approach kept a high true positive rate while reducing the false positive rate by around 50%. Also, the method can efficiently handle imbalanced datasets which is crucial for transactional monitoring. Further studies could be done using a larger dataset to analyse how the EM approach performs.

An auto-regressive (AR) outlier-based money laundering detection method is proposed by (Kannan and Somasundaram, 2017), called AROMLD. A regression deviation ranking is calculated for every transaction. Then a threshold limit is calculated by multiplying the interquartile range with a constant value (1.75 or 1.5). If the regression deviation ranking is larger than the threshold limit, the transaction is classed as an anomaly. The goal of this study is to reduce the computational complexity and data dimensionality for transactional monitoring. After comparing the results to other techniques, AROMLD had a shorter computational time, better accuracy, and ROC-AUC. Further research could be done using real data as synthetic data was used during this experiment. This could give a better representation of the model's performance.

2.7 Results: Evaluation Methods Used in Transaction Monitoring

Evaluation of transactional monitoring methods is crucial for potential users to assess the performance and establish the suitability of the method depending on their requirements. This section contains a discussion on the level of evaluation of the papers set by the author. An overview of all the evaluation techniques and the true positive/false positive rates (TPR/FPR) is analysed. Generally, evaluating unsupervised approaches is more challenging than evaluating supervised ones. In supervised methods, the presence of labelled data enables the use of mathematical metrics (i.e., TPR/FPR) to quantify performance. In contrast, unsupervised methods typically rely on other assessment techniques, such as analysing clusters and relationships, making objective evaluation less straightforward. However, several studies in this review applied unsupervised approaches while leveraging labelled data for evaluation, effectively assessing their models using supervised metrics. For example, (Zhiyuan et al., 2021) demonstrates this approach by incorporating labelled data to validate the models' outcomes. Evaluation of the method alone is not enough to determine the success and possible deployment of the approach. This is because some studies' results are low in quality or have misleading metrics. There's a need to evaluate the quality of the entire method, the dataset, and if it reaches the requirements (i.e., scalability, execution time, interpretability). The distribution of studies regarding their level rating is presented in Figure 2.6. The data was extracted by levelling the studies from one to six based on the criteria stated in Section 3. All the papers provide some sort of evaluation; hence, 0 papers are included in level one. 28% of the studies fell into levels two and three as they did not conduct any mathematical evaluations. A large number of studies only relied on demonstration or discussion/analysis to evaluate their methods. This is partially a result of authors being unable to acquire labelled data and utilising unsupervised approaches.

The majority of the papers are in level four, with levels four and five containing 62% of the studies. A small portion of papers conducted further analysis (level 6), as presented in Figure 2.6. To improve the evaluation of methods, papers should conduct further analysis, such as comparing the proposed method to existing rule-based methods or assessing if the method meets users' requirements. However, conducting further analysis will likely require cooperation with industry partners, which could be challenging to accomplish and may explain why only a few papers have done so. Gathering data from industrial partners will also be difficult, as the information will include confidential details that need to be anonymised.

Figure 2.7 presents the number of times each evaluation technique was applied in the reviewed studies. Most papers that conducted a mathematical analysis to calculate error rates used a combination of the following techniques: accuracy, recall, F1-score, and precision. Many authors also produced an ROC-AUC graph to give a visual representation of the trade-off between sensitivity and specificity. Recall is a measure of correctly detecting true positives (i.e., suspicious transactions that are predicted as suspicious).

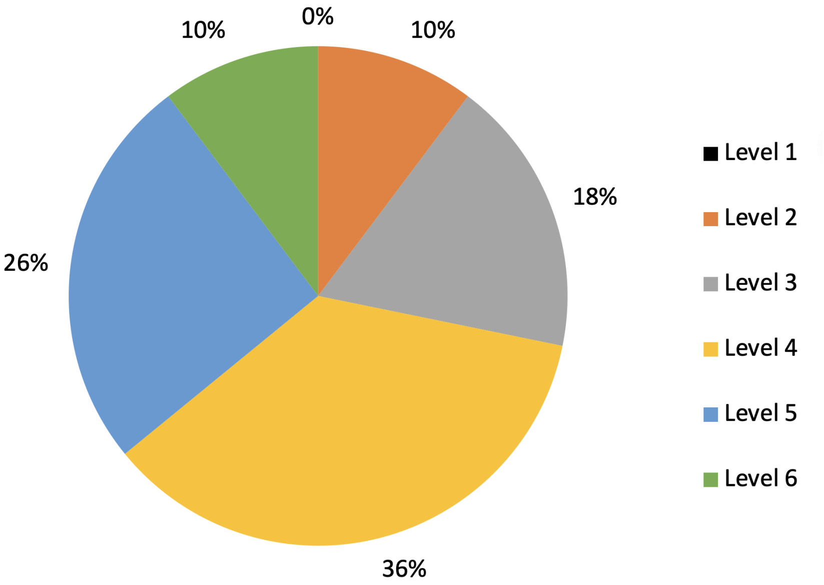


Figure 2.6: Evaluation level of studies that conducted experiments on transaction monitoring.

Precision is used to measure the rate of correctly classified suspicious transactions relative to all transactions labelled as suspicious. To get an overview of the correctness of the labelled transactions, accuracy is utilised. However, the accuracy rate can be misleading due to the imbalanced nature of transactional monitoring datasets. The F1 score, which considers both precision and recall, is often used to summarise the model's overall performance. Many studies also calculate the false positive rate (FPR). FPR is especially important in transactional monitoring, as current industry approaches have high rates, and minimising it is a major objective.

Papers that calculated the TPR and FPR for their approaches are presented in Table 2.4. A high TPR is essential for industry adoption, as institutions risk large fines if fraudulent transactions go undetected, while a low FPR enhances efficiency and reduces operational costs. However, interpreting and comparing these results directly is challenging due to substantial variations in dataset types, sizes, and quality across studies. These dataset inconsistencies impact performance and highlight key limitations within current research, motivating the need for standardised benchmarks. Although results are promising, further experimentation, particularly using standardised datasets, is necessary to strengthen evidence regarding method reliability and effectiveness. Additionally, methods evaluated solely on synthetic data should be tested with real transaction data to validate their practical applicability.

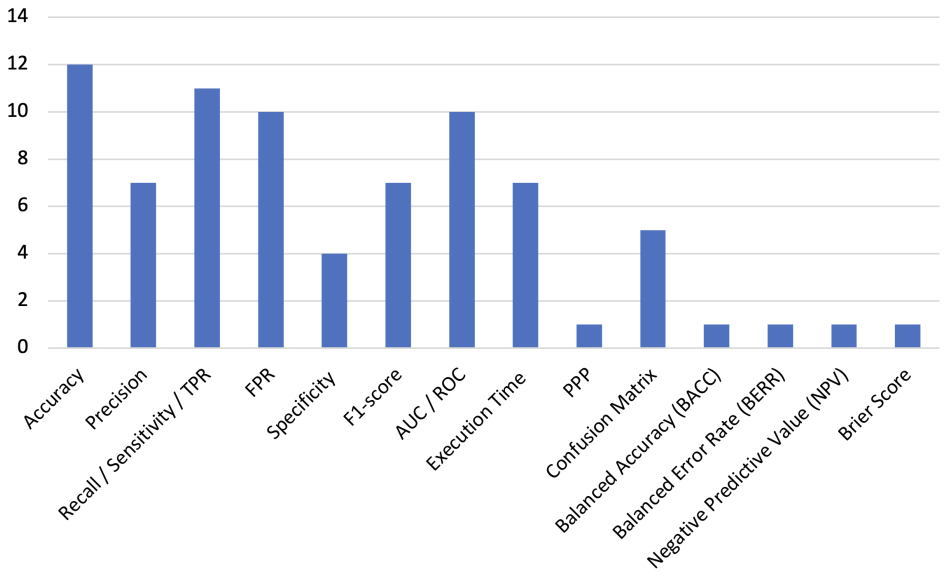


Figure 2.7: Most utilized evaluation techniques.

Table 2.4: Summary of Machine Learning Styles and Performance

Machine learning style	Paper	TPR	FPR
Unsupervised			
	Zhiyuan et al., 2021	100%	7%
	Alshantti and Rasheed, 2021	85.60%	16.20%
	J. D. J. Rocha-Salazar, M. J. Segovia-Vargas, and M. D. M. Camacho-Miñano, 2021	100.00%	13.60%
	Stojanović et al., 2021	92.70%	-
	Stojanović et al., 2021	93.20%	-
	Stojanović et al., 2021	98.50%	-
	Z. Chen, Khoa, et al., 2014	93.00%	1.60%
	Wang and Dong, 2009	72%	-
Supervised			
	Tundis, Nemalikanti, and Mühlhäuser, 2021	97.20%	6.70%
	Magomedov et al., 2018	-	4.20%
	Jun and Jian, 2005	-	3.40%
	Keyan and Tingting, 2011	79%	5.80%
	Lv, Ji, and J. L. Zhang, 2008	100%	76%
Ensemble			
	Stojanović et al., 2021	81.60%	-
	Stojanović et al., 2021	88.60%	-
	Stojanović et al., 2021	99.30%	-
Semi-supervised			
	Khac and Kechadi, 2010	100%	-

2.8 Key Shortcomings in Transaction Monitoring

It is evident that machine learning approaches produce promising results and can improve upon the current rule-based methods. However, further research and experiments are required before they can be used practically and deployed in the industry. Research should consider the shortcomings of the existing literature listed below when conducting future research.

- **Feature selection:** Several studies lack clarity in their feature selection process and do not explore the potential impact on results, which is crucial for decision-makers and regulators.

- **Datasets:** The use of diverse datasets across studies hinders the comparison of results between different approaches, highlighting the need for a standardised dataset to support future research and enhance machine learning methods.

- **Transaction types in Datasets:** The datasets lack sufficient attention regarding transaction types (e.g. cross border, SEPA, and cash transactions), which play a critical role in the risk-based monitoring approach required by authorities.

- **Dataset quality and evaluation:** Some studies are unreliable due to a lack of quality assurance and insufficient information regarding the dataset used, leading to ambiguous results. For instance, utilising a small dataset or the failure to disclose critical details, such as the number of normal and money laundering transactions, may oversimplify scenarios and produce skewed outcomes.

- **Evaluation methodology:** Most publications rely on a single dataset for evaluation, reducing reliability, and the metrics employed may be misleading.

- **Reproducibility:** The proposed methods are often difficult to reproduce due to the absence of detailed pre-processing explanations and the unavailability of datasets used in experiments.

- **Industrial collaboration:** Few studies have cooperated with institutions to understand their requirements and compare their transaction monitoring approaches. This may result from the difficulty and time constraints involved in finding suitable institutions for collaboration, leading to a gap between academia and industry and hindering the production of realistic methods and results.

2.9 Future Research Directions

This section outlines several areas that hold promise for further research and improvement in the field of transaction monitoring. Below, key research directions are proposed that aim to refine these methodologies and expand their application in the banking and financial sectors:

- **Machine Learning Methods:** Unsupervised learning methods have yet to be fully explored for transaction monitoring and could be the subject of future research. These methods offer benefits as they are more adaptable to changes in customer behaviour and the difficulties associated with acquiring accurately labelled data. Furthermore, the use of reinforcement learning, ensemble learning, and deep learning has been largely overlooked in transaction monitoring and could be explored further. Additionally, exploring transformer models could provide new insights and improvements in detecting complex patterns in transactional data.

- **Machine Learning Approaches:** Graph analysis is a promising approach that has yet to be fully explored in the field of transaction monitoring. By analysing social links and connections between individuals and groups, graph analysis has the potential to identify suspicious transactions and detect patterns and relationships that are indicative of money laundering. Additionally, future research could investigate the potential benefits of combining multiple machine learning approaches and aggregating the results for improved outcomes.

- **Dataset Consideration:** Studies in the field have been analysing transactional data mainly. However, future research could benefit from incorporating additional data sources such as KYC data, Company House data, or social media data to enhance the accuracy of transaction monitoring. This broader scope would offer a more holistic view of customer activities and relationships, enabling more robust and insightful risk assessments.

- **Dataset:** To improve the accuracy of transaction monitoring, future studies could concentrate on high-risk transaction types such as cross border, SEPA, and cash transactions. By doing so, the most significant transactions can be monitored with increased accuracy and a reduction in false positive results.

- **Data Creation:** The evaluation and comparison of different machine learning algorithms are complicated by the use of varying datasets. To address this issue, it is proposed that a publicly available, high-quality dataset that includes money laundering scenarios be created for researchers to use as a benchmark for evaluating their results.

- **Explainable AI:** The author would like to highlight that the development of explainable AI methods for transaction monitoring is a key direction for future research. This approach could facilitate greater industry adoption of complex techniques, address regulatory requirements for model transparency, and provide clarity on the decision-making processes and reasoning behind results. Based on the authors' review, there is currently very limited research in this area.

2.10 Summary

Anti-money laundering is conducted by financial institutions to prevent and reduce financial offences as well as to comply with authorities' legal obligations. Transactional monitoring is one type of AML approach done alongside others (i.e., CDD) which can help identify and report money launderers. Current monitoring methods within the industry are costly, time-consuming, and inefficient (high false positives), hence, enhancements are needed. However, producing an effective transactional monitoring method is demanding due to the increasing number of transactions, technology advancements, lack of ground truths, and unavailability of data. A review of transactional monitoring methods within the literature is conducted in this research by analysing different aspects of the approaches. First, the various types of pre-processing methods are discussed followed by an analysis of the datasets and feature selection techniques. Machine learning algorithms employed to create the transactional monitoring methods are discussed and reviewed next. Finally, the studies in the review are given an evaluation ranking and the types of evaluation techniques for the produced methods are examined.

Based on this research, the following shortcomings are identified. Regarding the datasets, there is a lack of quality assurance, insufficient attention to high-risk transaction types (e.g., cross-border), and a need for a base dataset to enable accurate comparisons of results. The evaluation of current approaches is generally done on a single dataset and could be examined further to increase reliability and trust. The reproducibility of studies is also problematic, reducing their reliability. Financial institutions evaluate large datasets; hence, efficient pre-processing techniques are required. Although research is being conducted on developing transactional monitoring methods, further research is necessary. Future research should consider the shortcomings of existing methods when producing new solutions. Less research has been done on graph analysis and deep learning approaches, which can be explored further. Different types of data, such as KYC data, can be used alongside transactional data. A potential research direction for more accurate and realistic results is to cooperate with financial institutions to produce transaction monitoring methods. This can reduce the academic-industry gap; however, it will be difficult to achieve as transactional data is highly sensitive.

For future work, the author will be attempting to collect a real dataset and identify the requirements of a transaction monitoring method for institutions. Then a method to detect suspicious transactions will be produced and evaluated that reaches the requirements identified. Reducing the number of alerted false positives compared to rule-based methods will be a major objective as this is a primary concern in the AML domain.

Chapter 2 References

Alshantti, Abdallah and Adil Rasheed (2021). “Self-Organising Map Based Framework for Investigating Accounts Suspected of Money Laundering”. In: *Frontiers in Artificial Intelligence* 4, pp. 1–15. ISSN: 2624-8212. DOI: 10.3389/frai.2021.761925.

Baldi, Pierre (July 2012). “Autoencoders, Unsupervised Learning, and Deep Architectures”. In: *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*. Vol. 27. Proceedings of Machine Learning Research. PMLR, pp. 37–49.

Breiman, Leo (2001). “Random Forests”. In: *Machine Learning* 45, pp. 5–32. DOI: 10.1023/A:1010933404324.

Breunig, Markus M. et al. (2000). “LOF: Identifying Density-Based Local Outliers”. In: *SIGMOD Rec.* 29.2, pp. 93–104. DOI: 10.1145/335191.335388.

Butgereit, L. (2021). “Anti Money Laundering: Rule-Based Methods to Identify Funnel Accounts”. In: *2021 Conference on Information Communications Technology and Society (ICTAS)*, pp. 21–26. DOI: 10.1109/ICTAS50802.2021.9394990.

Cai, Jie et al. (2018). “Feature selection in machine learning: A new perspective”. In: *Neurocomputing* 300, pp. 70–79.

Camino, R. D. et al. (2017). "Finding Suspicious Activities in Financial Transactions and Distributed Ledgers". In: *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 787–796. ISBN: 2375-9259. DOI: 10.1109/ICDMW.2017.109.

Canhoto, Ana Isabel (2021). "Leveraging machine learning in the global fight against money laundering and terrorism financing: An affordances perspective". In: *Journal of Business Research* 131, pp. 441–452. ISSN: 0148-2963. DOI: 10.1016/j.jbusres.2020.10.012.

Chandrashekar, Girish and Ferat Sahin (2014). "A survey on feature selection methods". In: *Computers & Electrical Engineering* 40.1, pp. 16–28. ISSN: 0045-7906. DOI: 10.1016/j.compeleceng.2013.11.024.

Chang, Chih-Chung and Chih-Jen Lin (2011). "LIBSVM: A library for support vector machines". In: *ACM Transactions on Intelligent Systems and Technology* 2.3, pp. 1–27. ISSN: 2157-6904. DOI: 10.1145/1961189.1961199.

Chen, Gang and Jin Chen (2015). "A novel wrapper method for feature selection and its applications". In: *Neurocomputing* 159, pp. 219–226. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2015.01.070.

Chen, Z., L. Dinh Van Khoa, et al. (2014). "Exploration of the effectiveness of expectation maximization algorithm for suspicious transaction detection in anti-money laundering". In: *2014 IEEE Conference on Open Systems (ICOS)*, pp. 145–149. DOI: 10.1109/ICOS.2014.7042645.

Chen, Z., L. D. Van Khoa, E. N. Teoh, et al. (2018). "Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review". In: *Knowledge and Information Systems* 57, pp. 245–285. DOI: 10.1007/s10115-017-1144-z.

Cherrington, M. et al. (2019). "Feature Selection: Filter Methods Performance Challenges". In: *2019 International Conference on Computer and Information Sciences (ICCIS)*, pp. 1–4. DOI: 10.1109/ICCISci.2019.8716478.

Cortes, C. and V. Vapnik (1995). "Support-Vector Networks". In: *Machine Learning* 20.3, pp. 273–297. DOI: 10.1023/A:1022627411411.

Desrousseaux, R., G. Bernard, and J. J. Mariage (2021). "Profiling Money Laundering with Neural Networks: a Case Study on Environmental Crime Detection". In: *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 364–369. ISBN: 2375-0197. DOI: 10.1109/ICTAI52525.2021.00059.

Eifrem, Emil (2019). "How graph technology can map patterns to mitigate money-laundering risk". In: *Computer Fraud & Security* 2019.10, pp. 6–8. DOI: 10.1016/S1361-3723(19)30105-8.

Gao, H. et al. (2022). "TSMAE: A Novel Anomaly Detection Approach for Internet of Things Time Series Data Using Memory-Augmented Autoencoder". In: *IEEE Transactions on Network Science and Engineering*. ISSN: 2327-4697. DOI: 10.1109/TNSE.2022.3163144.

Gao, Z. (2009). "Application of Cluster-Based Local Outlier Factor Algorithm in Anti-Money Laundering". In: *2009 International Conference on Management and Service Science*, pp. 1–4. DOI: 10.1109/ICMSS.2009.5302396.

Guevara, Jorge, Olmer Garcia-Bedoya, and Oscar Granados (2020). "Machine Learning Methodologies Against Money Laundering in Non-Banking Correspondents". In: *Applied Informatics*. Springer International Publishing, pp. 72–88. DOI: doi.org/10.1007/978-3-030-61702-8_6.

Hayble-Gomes, E. (2023). “The use of predictive modeling to identify relevant features for suspicious activity reporting”. In: *Journal of Money Laundering Control* 26.4, pp. 806–830. DOI: 10.1108/JMLC-02-2022-0034.

Helmy, Tamer H. et al. (2016). “Design of a monitor for detecting money laundering and terrorist financing”. In: *Journal of Theoretical and Applied Information Technology* 85.3, pp. 425–436.

Huang, D. et al. (2018). “CoDetect: Financial Fraud Detection With Anomaly Feature Detection”. In: *IEEE Access* 6, pp. 19161–19174. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2018.2816564.

Hussain, Walayat, José M. Merigó, and Muhammad Raheel Raza (2022). “Predictive intelligence using ANFIS-induced OWAWA for complex stock market prediction”. In: *International Journal of Intelligent Systems* 37.8, pp. 4586–4611. ISSN: 0884-8173. DOI: 10.1002/int.22732.

Jamshidi, M. B. et al. (2019). “A novel multiobjective approach for detecting money laundering with a neuro-fuzzy technique”. In: *IEEE 16th International Conference on Networking, Sensing and Control (ICNSC)*, pp. 454–458. DOI: 10.1109/ICNSC.2019.8743234.

Jolly, Jasper (2020). *Commerzbank fined by UK watchdog FCA for money laundering failings.* The Guardian. Available from: https://www.theguardian.com/business/2020/jun/17/commerzbank-fined-uk-watchdog-fca-money-laundering-failings [Accessed 4 Apr. 2023].

Jullum, Martin et al. (2020). “Detecting money laundering transactions with machine learning”. In: *Journal of Money Laundering Control* 23.1, pp. 173–186. ISSN: 1368-5201. DOI: 10.1108/JMLC-07-2019-0055.

Jun, Tang and Yin Jian (2005). “Developing an intelligent data discriminating system of anti-money laundering based on SVM”. In: *2005 International Conference on Machine Learning and Cybernetics*. Vol. 6, pp. 3453–3457. ISBN: 2160-1348. DOI: 10.1109/ICMLC.2005.1527539.

Kannan, S. and K. Somasundaram (2017). “Autoregressive-based outlier algorithm to detect money laundering activities”. In: *Journal of Money Laundering Control* 20.2, pp. 190–202. ISSN: 1368-5201. DOI: 10.1108/JMLC-07-2016-0031.

Karim, Md. Rezaul et al. (2024). “Scalable Semi-Supervised Graph Learning Techniques for Anti Money Laundering”. In: *IEEE Access* 12, pp. 50012–50029. DOI: 10.1109/ACCESS.2024.3383784.

Keele, Staffs (2007). *Guidelines for performing systematic literature reviews in software engineering.* Report. Technical report, Ver. 2.3 EBSE Technical Report. EBSE.

Ketenci, Utku Gorkem et al. (2021). “A Time-Frequency Based Suspicious Activity Detection for Anti-Money Laundering”. In: *IEEE Access* 9, pp. 59957–59967. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3072114.

Keyan, L. and Y. Tingting (2011). “An Improved Support-Vector Network Model for Anti-Money Laundering”. In: *2011 Fifth International Conference on Management of e-Commerce and e-Government*, pp. 193–196. DOI: 10.1109/ICMeCG.2011.50.

Khac, N. A. Le and M. Kechadi (2010). “Application of Data Mining for Anti-money Laundering Detection: A Case Study”. In: *IEEE International Conference on Data Mining Workshops*, pp. 577–584. ISBN: 2375-9259. DOI: 10.1109/ICDMW.2010.66.

Koo, Kyungmo, Minyoung Park, and Byungun Yoon (2024). “A Suspicious Financial Transaction Detection Model Using Autoencoder and Risk-Based Approach”. In: *IEEE Access* 12, pp. 68926–68939. DOI: 10.1109/ACCESS.2024.3399824.

Kute, Dattatray Vishnu et al. (2021). “Deep Learning and Explainable Artificial Intelligence Techniques Applied for Detecting Money Laundering–A Critical Review”. In: *IEEE Access* 9, pp. 82300–82317. DOI: 10.1109/ACCESS.2021.3086230.

Larik, A. S. and S. Haider (2011). “Clustering based anomalous transaction reporting”. In: *Procedia Computer Science*. Vol. 3, pp. 606–610. DOI: 10.1016/j.procs.2010.12.101.

Leite, Gleidson Sobreira, Adriano Bessa Albuquerque, and Plácido Rogerio Pinheiro (2019). “Application of Technological Solutions in the Fight against Money Laundering - A Systematic Literature Review”. In: *Applied Sciences* 9.22, pp. 1–29. DOI: 10.3390/app9224800.

Liu, F. T., K. M. Ting, and Z. Zhou (2008). “Isolation Forest”. In: *Eighth IEEE International Conference on Data Mining*, pp. 413–422. ISBN: 2374-8486. DOI: 10.1109/ICDM.2008.17.

Liu, Xuan and Pengzhu Zhang (2010). “A Scan Statistics Based Suspicious Transactions Detection Model for Anti-money Laundering (AML) in Financial Institutions”. In: *International Conference on Multimedia Communications*, pp. 210–213. ISBN: 978-0-7695-4136-5. DOI: 10.1109/MEDIACOM.2010.37.

Lv, L. T., N. Ji, and J. L. Zhang (2008). “A RBF neural network model for anti-money laundering”. In: *International Conference on Wavelet Analysis and Pattern Recognition*. Vol. 1, pp. 209–215. ISBN: 2158-5709. DOI: 10.1109/ICWAPR.2008.4635778.

Magomedov, S. et al. (2018). “Anomaly detection with machine learning and graph databases in fraud management”. In: *International Journal of Advanced Computer Science and Applications(IJACSA)* 9.11, pp. 33–38. DOI: 10.14569/IJACSA.2018.091104.

McDowell, John and Gary Novis (2001). “The consequences of money laundering and financial crime”. In: *Economic Perspectives* 6.2, pp. 6–10.

Michalak, K. and J. Korczak (2011). “Graph mining approach to suspicious transaction detection”. In: *Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp. 69–75.

Naheem, Mohammed Ahmad (2016). “Money laundering: A primer for banking staff”. In: *International Journal of Disclosure and Governance* 13.2, pp. 135–156. ISSN: 1746-6539.

Pargent, Florian et al. (2022). “Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features”. In: *Computational Statistics*, pp. 1–22. ISSN: 1613-9658. DOI: 10.1007/s00180-022-01207-6.

Phyu, The Hnin and Surapong Uttama (2023). “Improving Classification Performance of Money Laundering Transactions Using Typological Features”. In: *2023 7th International Conference on Information Technology (InCIT)*, pp. 520–525. DOI: 10.1109/InCIT60207.2023.10413155.

Prado, F. F. and L. A. Digiampietri (2020). “A systematic review of automated feature engineering solutions in machine learning problems”. In: *XVI Brazilian Symposium on Information Systems (SBSI’20)*, pp. 1–7. DOI: 10.1145/3411564.3411610.

Raza, Saleha and Sajjad Haider (2011). “Suspicious activity reporting using dynamic bayesian networks”. In: *Procedia Computer Science* 3, pp. 987–991. ISSN: 1877-0509. DOI: 10.1016/j.procs.2010.12.162.

Rocha-Salazar, J. D. J., M. J. Segovia-Vargas, and M. D. M. Camacho-Miñano (2021). “Money laundering and terrorism financing detection using neural networks and an abnormality indicator”. In: *Expert Systems with Applications* 169, pp. 1–15. DOI: 10.1016/j.eswa.2020.114470.

Rouhollahi, Z. et al. (2021). “Towards Proactive Financial Crime and Fraud Detection through Artificial Intelligence and RegTech Technologies”. In: *The 23rd International Conference on Information Integration and Web Intelligence*, pp. 538–546. DOI: 10.1145/3487664.3487740.

Ruchay, Alexey et al. (2023). “The Imbalanced Classification of Fraudulent Bank Transactions Using Machine Learning”. In: *Mathematics* 11.13, p. 2862. DOI: 10.3390/math11132862.

Rui, Xu and D. Wunsch (2005). “Survey of clustering algorithms”. In: *IEEE Transactions on Neural Networks* 16.3, pp. 645–678. ISSN: 1941-0093. DOI: 10.1109/TNN.2005.845141.

Saar-Tsechansky, M. and F. Provost (2007). “Handling missing values when applying classification models”. In: *Journal of Machine Learning Research* 8, pp. 1625–1657.

Schmidhuber, Jürgen (2015). “Deep learning in neural networks: An overview”. In: *Neural Networks* 61, pp. 85–117. ISSN: 0893-6080. DOI: 10.1016/j.neunet.2014.09.003.

Shokry, Amr Ehab Muhammed, Mohammed Abo Rizka, and Nevine Makram Labib (2020). “Counter terrorism finance by detecting money laundering hidden networks using unsupervised machine learning algorithm”. In: *International Conference on e-Learning*, pp. 89–97. DOI: 10.33965/ict_csc_wbc_2020_2020081012.

Stojanović, Branka et al. (2021). “Follow the Trail: Machine Learning for Fraud Detection in Fintech Applications”. In: *Sensors* 21.5. ISSN: 1424-8220. DOI: 10.3390/s21051594.

Tai, C. and T. Kan (2019). “Identifying Money Laundering Accounts”. In: *International Conference on System Science and Engineering (ICSSE)*, pp. 379–382. ISBN: 2325-0925. DOI: 10.1109/ICSSE.2019.8823264.

Tundis, A., S. Nemalikanti, and M. Mühlhäuser (2021). “Fighting organized crime by automatically detecting money laundering-related financial transactions”. In: *The 16th International Conference on Availability, Reliability and Security*. Vol. 38, pp. 1–10. DOI: 10.1145/3465481.3469196.

Usman, A., N. Naveed, and S. Munawar (2023). “Intelligent Anti-Money Laundering Fraud Control Using Graph-Based Machine Learning Model for the Financial Domain”. In: *Journal of Cases on Information Technology (JCIT)* 25.1, pp. 1–20. DOI: 10.4018/JCIT.316665.

Wang, X. and G. Dong (2009). “Research on Money Laundering Detection Based on Improved Minimum Spanning Tree Clustering and Its Application”. In: *Second International Symposium on Knowledge Acquisition and Modeling*. Vol. 2, pp. 62–64. DOI: 10.1109/KAM.2009.221.

Xia, Pingfan et al. (2024). “A Novel Heuristic-Based Selective Ensemble Prediction Method for Digital Financial Fraud Risk”. In: *IEEE Transactions on Engineering Management* 71, pp. 8002–8018. DOI: 10.1109/TEM.2024.3385298.

Zhang, S., X. Wu, and M. Zhu (2010). “Efficient missing data imputation for supervised learning”. In: *9th IEEE International Conference on Cognitive Informatics (ICCI’10)*, pp. 672–679. DOI: 10.1109/COGINF.2010.5599826.

Zhang, Y. and P. Trubey (2019). “Machine Learning and Sampling Scheme: An Empirical Study of Money Laundering Detection”. In: *Comput Economics* 54, pp. 1043–1063. DOI: 10.1007/s10614-018-9864-z.

Zhiyuan, Chen et al. (2021). “Variational Autoencoders and Wasserstein Generative Adversarial Networks for Improving the Anti-Money Laundering Process”. In: *IEEE Access* 9, pp. 83762–83785. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3086359.

Chapter 3

Semi-Structured Interviews: Industry view of AML

Version of record available at: B. Oztas, D. Cetinkaya, F. Adedoyin, M. Budka, H. Dogan and G. Aksu, ”Transaction monitoring in anti-money laundering: A qualitative analysis and points of view from industry,” Future Generation Computer Systems, vol. 159, 2024, pp. 161-171. ISSN 0167-739X.
DOI: https://doi.org/10.1016/j.future.2024.05.027
Licence: CC BY-NC 4.0

Version of record available at: B. Oztas, D. Cetinkaya, F. Adedoyin, M. Budka, H. Dogan, and G. Aksu. ”Perspectives from Experts on Developing Transaction Monitoring Methods for Anti-Money Laundering,” in 2023 IEEE International Conference on e-Business Engineering (ICEBE), Sydney, Australia, 2023, pp. 39-46.
DOI: 10.1109/ICEBE59045.2023.00024
Licence: © 2023 IEEE.

3.1 Introduction

This chapter builds on the lack of academic-industrial collaboration identified in Chapter 2, addressing objective (ii): Analysing the insights and expectations of AML specialists. Figure 3.1 illustrates the problem and the input from Chapter 2 that informed the development of Chapter 3. This chapter has been published as a journal article in the *Future Generation Computer Systems* journal as part of the integrated thesis. Additionally, a long paper has been published in the *2023 IEEE International Conference on e-Business Engineering*, which won the best student paper award.

As financial institutions continue to face significant challenges in transaction monitoring, including high false positive rates, evolving money laundering scenarios, and the need for an efficient and effective detection process (Eifrem, 2019). Further research is essential to improve transaction monitoring methods, enhance crime prevention, and alleviate these challenges for financial institutions.

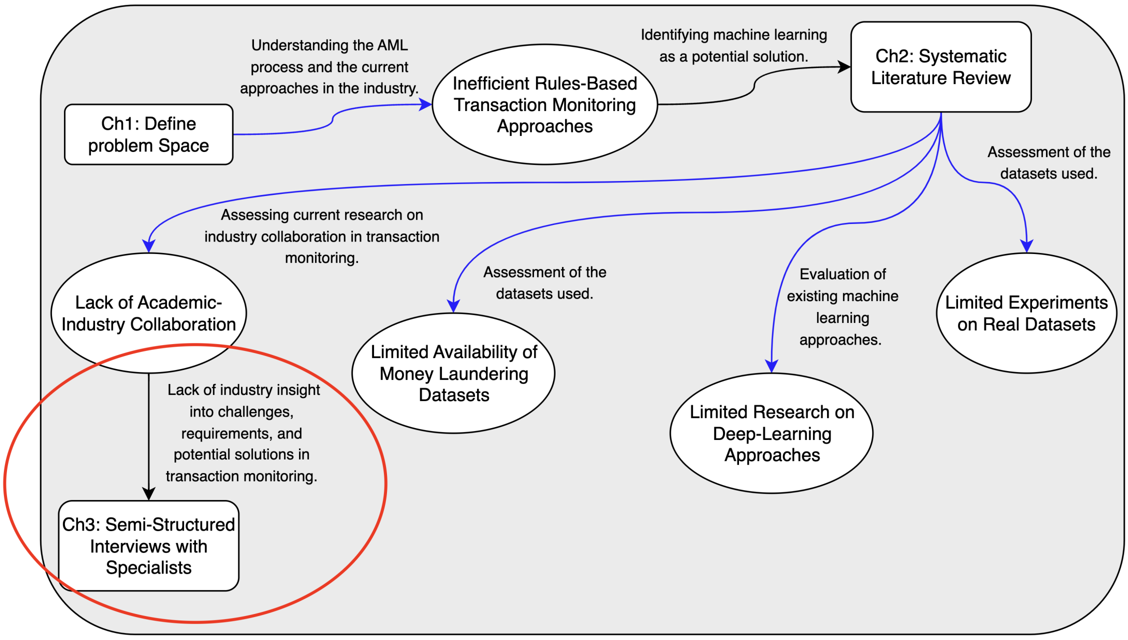

Ch1: Define problem Space
Understanding the AML process and the current approaches in the industry.
Inefficient Rules-Based Transaction Monitoring Approaches
Identifying machine learning as a potential solution.
Ch2: Systematic Literature Review
Assessment of the datasets used.
Assessing current research on industry collaboration in transaction monitoring.
Lack of Academic-Industry Collaboration
Assessment of the datasets used.
Limited Availability of Money Laundering Datasets
Evaluation of existing machine learning approaches.
Limited Research on Deep-Learning Approaches
Limited Experiments on Real Datasets
Lack of industry insight into challenges, requirements, and potential solutions in transaction monitoring.
Ch3: Semi-Structured Interviews with Specialists

Figure 3.1: Overview of the PhD (extracted from Figure 1.3)

This research aims to aid researchers and practitioners in the AML domain to develop improved and more efficient transaction monitoring methods. To comprehensively understand the field and achieve this aim, three objectives are established: 1. To identify the issues and challenges faced by financial institutions during the transaction monitoring process, 2. to understand the requirements a new transaction monitoring method requires to be adoptable and successful in the industry, and 3. to gather specialists' opinions on potential solutions and the future of the transaction monitoring domain. Semi-structured interviews with AML experts are conducted and then analysed during this chapter. The current literature lacks industrial collaboration (Kute et al., 2021) with limited studies analysing the industry's problems surrounding transaction monitoring and the requirements a new method needs to meet the industry's demands. This study expands upon the work presented by B. Oztas et al., 2023b and contributes to the existing literature by collaborating with AML specialists in the industry to gain a better understanding of the field.

An analysis of existing literature related to the study's objectives is presented in Section 3.2. Section 3.3 explains the methodology that was employed to conduct this research including the use of semi-structured interviews, data analysis, and limitations. The findings and analysis of them are presented in Section 3.4. Finally, the summary is presented in Section 3.5 summarising the chapter.

3.2 Literature Review

This review explores the current literature on the problems and challenges within financial institutions around transaction monitoring, the requirements for new transaction monitoring methods to be adopted and successful in the industry, and future trends and prospects of transaction monitoring. Key problems institutions face and need to resolve include regulatory challenges, data quality issues, high false positives, operational and implementation costs.

These issues need to be addressed by financial institutions to effectively detect money launderers. The highlighted requirements in the literature consist of having a scalable approach, reduction in false positives, compliance with regulator's standards, attaining higher accuracy, and flexibility in the method. Future research should consider these requirements when developing a new approach. The literature suggests and is optimistic about the use and adoption of artificial intelligence and machine learning for transaction monitoring. Information sharing and big data analytics are two additional areas that are identified in the literature as future trends and prospects for transaction monitoring.

Financial institutions must comply with the AML regulations. However, institutions find it challenging to keep up and maintain compliance with the constantly evolving regulatory landscape and lack of international standards in transaction monitoring (Dill, 2021). Poor data quality has also been mentioned as a challenge for transaction monitoring in the literature. Inaccurate and incomplete data can impede the transaction monitoring methods' effectiveness (Halter et al., 2011). Data issues can result in high false positive rates, impacting the institution's ability to detect suspicious activity. The high rates of false positives generated are a major problem for financial institutions. The false positive alerts lead to a vast amount of investigations which increase the institution's costs, putting pressure on banks to improve the efficiency and effectiveness of their monitoring methods (Pontes et al., 2022). Implementation challenges of new transaction monitoring methods are also identified in the literature. Data privacy and protection is one area that makes implementation difficult as systems require the collection and sharing of personal and financial data when implementing a new method (X. Liu and P. Zhang, 2008). Additionally, the rapid pace of technological advancements along with banks' older technological systems makes it difficult to adapt and implement new solutions effectively (Zavoli and King, 2021). The cost of implementing a new monitoring system will also make it challenging for institutions, especially for smaller institutions with limited budgets and resources (Simonova, 2011; Mekpor, 2019). In addition, the changing of criminal techniques was identified as a challenge for institutions as the current monitoring methods struggle to keep up with the changes. The existing transaction monitoring methods need to be continuously updated and improved (e.g. creating new rules and scenarios) which is complex and costly to the organisation (Simonova, 2011). Another challenge identified is monitoring cross-border transactions due to the varying regulatory requirements and standards across jurisdictions (X. Liu and P. Zhang, 2008). Effective monitoring requires collaboration between financial institutions and regulators (Viritha, Mariappan, and Venkatachalapathy, 2015).

Accuracy is a major area that transaction monitoring methods need to improve and therefore a key requirement. Improving the detection accuracy will be a crucial element when measuring the adoptability and success of a new transaction monitoring method (Veyder, 2003). A system with higher accuracy can minimise the number of false positives generated, addressing a critical challenge that significantly impacts the effectiveness of transaction monitoring. Scalability is also a requirement for a new transaction monitoring method to handle the rising volume of data given the increasing size and complexity of transactions in financial institutions (Tertychnyi et al., 2020). A new method should be capable of horizontal and vertical scaling to ensure the approach can expand with demand (Ali, 2019). Adaptability to address new types of money laundering and terrorist financing should be implemented in a new transaction monitoring method to be deemed successful (Veyder, 2003). The requirement of adaptability can reduce regulatory pressures and prevent reputational damage. Another requirement for a transaction monitoring method to be adoptable is easy implementation into the institutions. The method must integrate with existing systems and require minimal training for end-users (Han, Y. Huang, S. Liu, et al., 2020). This will allow for a smooth transition from the current approaches to more sophisticated and enhanced methods. The cost-effectiveness of the method is also identified as a requirement, a positive return on investment should be shown to justify its implementation (Canhoto, 2021). A new method should reduce investigational and operational costs that arise from the large number of false positive alerts generated, which leads to hiring more employees (Berkan Oztas et al., 2022). Complying with regulations is another important requirement of a new transaction monitoring method. An institution must ensure that the method they implement is up to date with the regulations and must be continuously updated with the evolving regulatory requirements (Han, Y. Huang, S. Liu, et al., 2020). A risk-based approach is a regulatory requirement for a new method. The method should be customisable to address institution-specific risks and prioritise high-risk customers and industries (FATF, 2023).

The future of transaction monitoring will likely include artificial intelligence and machine learning. In the recent literature, the application of artificial intelligence and machine learning has gained considerable interest in transaction monitoring. Incorporating new and innovative solutions can improve the existing methods in the industry by increasing the accuracy and efficiency of detecting suspicious transactions (Z. Chen, Van Khoa, Teoh, et al., 2018). Many researchers argue that these innovative solutions will become a crucial component of the future of AML. The existing literature also discusses future AML efforts to collaborate and share information between financial institutions, law enforcement agencies, and regulators (Zolkaflil, Omar, and Syed Mustapha Nazri, 2019). Sharing data can lead to a higher accuracy rate and better detection of money launderers as institutions will have a clearer view of customer behaviour over multiple institutions. Sharing information and data can also lead to developing superior and standardised processes for transaction monitoring (Dehouck and Goede, 2021). The increasing availability of big data and advanced analytics is another area that is expected to affect the transaction monitoring domain in the future (Hasan, Popp, and Oláh, 2020). Institutions can leverage big data to enable the adopting of machine learning and enhance transaction monitoring solutions.

3.3 Research Methodology

This study adopts qualitative research by conducting semi-structured interviews with AML specialists. The overall aim of the study is to provide the field of research with a greater understanding of the transaction monitoring domain to aid in developing a new transaction monitoring approach. In this study, a qualitative approach is taken as a quantitative approach can be limiting in investigating the complexities of the transaction monitoring domain, such as understanding the contextual factors and nuances that may influence the development of a new approach (Wronka, 2022). Qualitative research can provide a deeper and more thorough insight into the opinions, knowledge, and experience of the transaction monitoring specialists (Dicicco-Bloom and Crabtree, 2006). Given that the research is exploratory in nature, utilising a qualitative research design will enable the gathered data to speak for itself. This study contributes to the literature by addressing the gap that currently exists between academia and industry in the transaction monitoring domain. Furthermore, it provides the requirements that a new transaction monitoring method needs to be useful and successful in the industry, from the specialists' points of view (Arora, Sabetzadeh, and Briand, 2019).

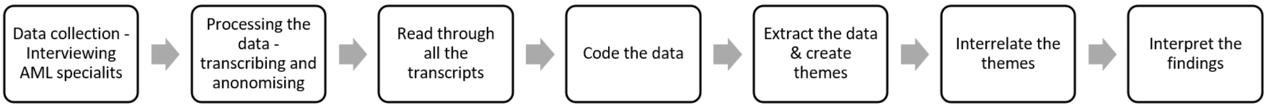

Data collection - Interviewing AML specialits
Processing the data - transcribing and anonomising
Read through all the transcripts
Code the data
Extract the data & create themes
Interrelate the themes
Interpret the findings

Figure 3.2: The Methodology Process.

3.3.1 Research Process

Despite the specialised nature of transaction monitoring and the challenges of finding individuals who meet the required criteria, purposive sampling was applied (Palinkas et al., 2015). Specifically, the authors identified specialists on LinkedIn who possessed at least nine years of transaction monitoring experience, ensuring a robust level of expertise. This method enabled the identification of professionals with the required level of expertise in transaction monitoring, who could provide valuable insights and perspectives on the subject matter. The selection of interviewees took into account their representativeness, determined by their expertise in the field of AML, as well as, their experience within the banking sector. This approach generated a group of specialists with the necessary characteristics to investigate the transaction monitoring domain to provide a greater understanding of the field and aid in developing a new transaction monitoring approach. The study's objectives were explained to potential participants and an information sheet outlining the study's parameters was provided. The 8 participants were required to complete and sign a participation agreement form to be included in the research. All relevant information and documentation were shared via email. Prior to the interview sessions, participants were asked to review the interview questionnaire to familiarise themselves with the questions (see Appendix A). The questions were developed through an analysis of the literature and collaboration with the industrial partner. All questions from the questionnaire were posed to every participant; however, additional questions were introduced during the interviews based on the direction each conversation took. During the interviews, participants were encouraged to express their perspectives and insights without any influence or bias. The saturation approach was employed to determine the sample size. New data, in the form of semi-structured interviews, was collected until no new information or insights were generated. Figure 3.2 presents the complete process of the methodology.

3.3.2 Data Collection Techniques

This study used an inductive research approach and therefore employed semi-structured interviews as the data collection technique (Thomas, 2006). The inductive research approach was chosen and utilised due to its capacity to examine the patterns and behaviours of domain experts, and to develop theories about the participants that connect to the chapter's objectives (Stevens, 2020). Semi-structured interviews were selected because they can be a valuable tool

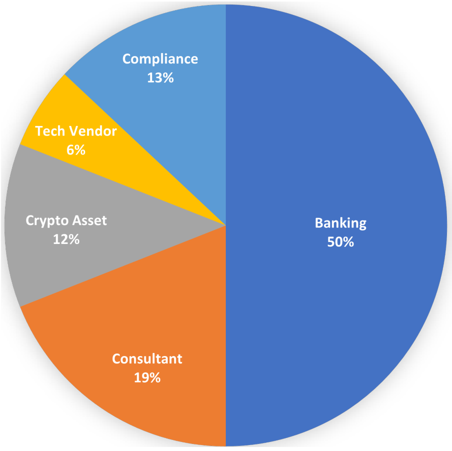


Figure 3.3: Interviewees experience in different domains.

Years of Experience	Total	Percentage
≤ 9	1	12.50%
10 - 19	5	62.5%
20 - 29	1	12.50%
30 - 39	1	12.50%
Total	8	100%

Table 3.1: Interviewees' years of experience in the transaction monitoring AML domain.

for researchers in the transaction monitoring domain, as they allow for an in-depth understanding of the participant's experiences, opinions, and knowledge related to the subject (Johnson and Kessler, 2019). Table 3.1 shows interviewees' years of experience in the AML transaction monitoring domain. This method allows for follow-up questions and exploration of areas in more detail, providing a more nuanced understanding. Participants can share their perspectives and provide insights that the researcher may not have previously considered, identifying potential blind spots and generating new ideas. The interviews lasted approximately 60 minutes and were conducted online. Once all the interviews were complete the data was transcribed and anonymised by the author for data analysis to then take place. The data collected was comprehensive while avoiding over-reliance on excessive sample size.

3.3.3 Data Analysis

This study refers to (J. W. Creswell and J. D. Creswell, 2018) for the data analysis process. The data analysis process began with an initial review of all the data, which provided an overall understanding of the information and an opportunity for reflection on its meaning, depth, and potential use. During this initial phase, I carefully read through the entire dataset while taking detailed annotations to capture first impressions. The next step involved coding, which entailed organising the data into labelled categories using the AML experts' language. The researchers adopted an inductive coding approach, a bottom-up method of data analysis that identifies patterns and themes from the data itself. This approach was selected as it allows for the exploration of new ideas and concepts that may not have been considered before and offers flexibility and adaptability in the analysis process while minimising researcher biases. Following the coding process, the researchers generated descriptions and themes, which were then used to develop complex analysis layers by interconnecting themes. Finally, the findings of the analysis were represented visually (Figure 3.4) and by a detailed discussion of several themes. Figure 3.3 highlights the interviewee's experience across various domains, with most specialists having experience in multiple industries. Table 3.1 presents the interviewee's years of experience within the AML domain.

3.3.4 Limitations

The study included a limited number of participants due to time constraints and the difficulty of acquiring participants with the relevant skills and knowledge. Hence, the data may not represent the entire AML and transaction monitoring domain which could limit the findings. Another limitation is the reliance on the author's interpretations and analysis of the data which can cause subjectivity in the findings. Future work should attempt to gather a larger number of specialist participants to gain a better understanding of the transaction monitoring domain. Also, more research is required to reach conclusive decisions on topics such as the requirements of a new transaction monitoring method as the author's perspectives have been presented.

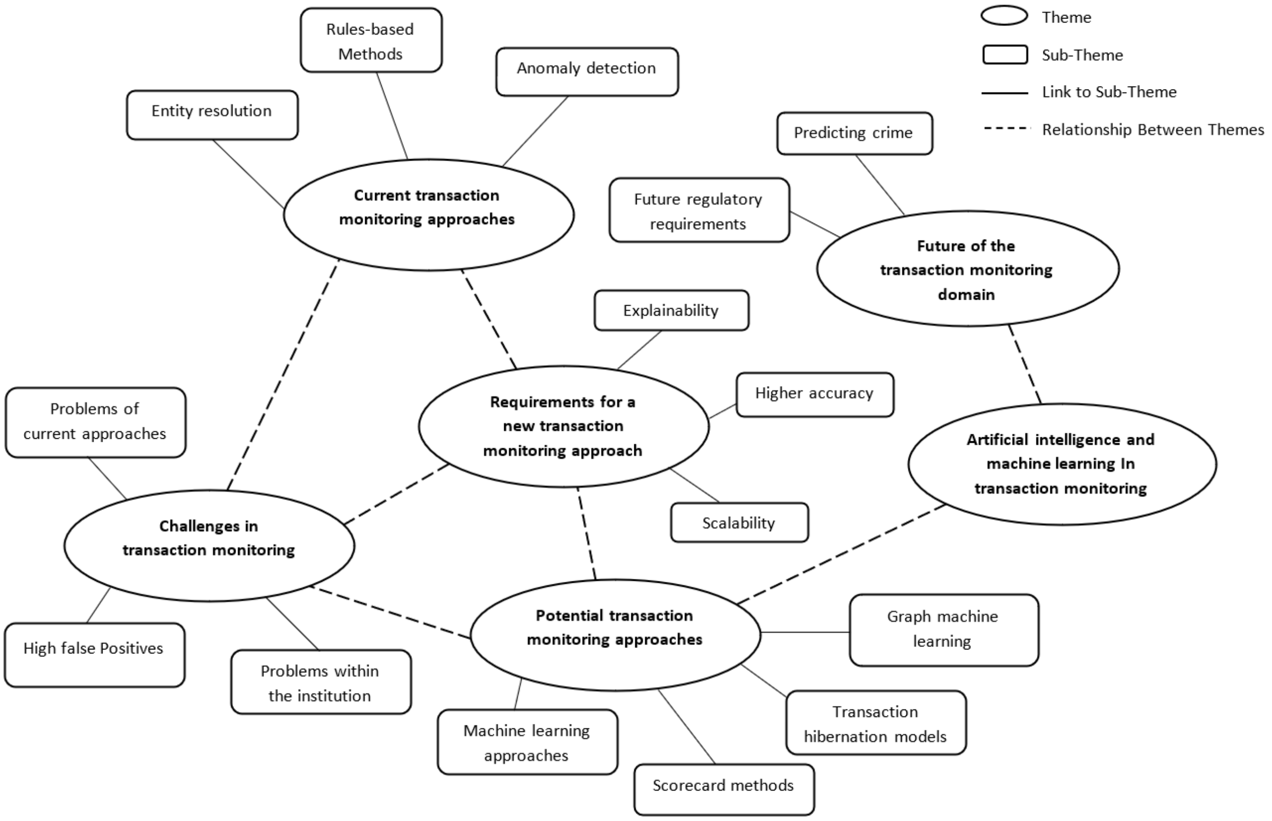

Theme
Sub-Theme
Link to Sub-Theme
Relationship Between Themes
Rules-based Methods
Anomaly detection
Entity resolution
Current transaction monitoring approaches
Predicting crime
Future regulatory requirements
Future of the transaction monitoring domain
Explainability
Higher accuracy
Requirements for a new transaction monitoring approach
Scalability
Problems of current approaches
Challenges in transaction monitoring
Artificial intelligence and machine learning In transaction monitoring
High false Positives
Problems within the institution
Potential transaction monitoring approaches
Graph machine learning
Transaction hibernation models
Machine learning approaches
Scorecard methods

Figure 3.4: Relationship between themes and sub-themes.

3.4 Findings and Discussion

This section presents the results of the semi-structured interviews conducted with AML specialists. The findings are themed into six areas: Current transaction monitoring approaches, Challenges in transaction monitoring, Artificial intelligence and machine learning in transaction

Table 3.2: Quotes of interviewees on the current transaction monitoring methods in the industry.

	What are the current transaction monitoring methods used in the industry?
1	*'So currently we heavily use rule-based systems, with scenario-based rules, where you alert if something is triggered'*
2	*"Mainly rule-based methods and a lot of statistical analysis is done to understand where it's appropriate to set thresholds"*
3	*"We detect anomalies in customer's behaviours at the transactional level. So we're looking for indications of financial crime using machine learning"*
4	*"A rule-based method, mostly a vendor solution. Generally, it's going to be one that has a library you can pick and choose from, ones which allow you to build rules"*
5	*"Most of our solution is based on rules but we do incorporate a bit of entity resolution and network analysis"*

monitoring, Requirements, Future of the domain, and Transaction monitoring approaches. A comprehensive analysis of the interviews, complemented by direct quotations from the specialists, is provided in tabular format. The tables also include the coding and thematic classification of the quotes, providing a clear and organised presentation of the data. An in-depth discussion and interpretation of the findings are conducted to better understand the implications of these insights for the transaction monitoring domain. The relationship between the different themes and various relevant sub-themes is presented in Figure 3.4.

3.4.1 Current Transaction Monitoring Approaches

The findings from the semi-structured interviews with transaction monitoring specialists suggest that currently, the industry is heavily reliant on rule-based systems for detecting money laundering and terrorist financing, which agrees with the current literature (Kumari and Gupta, 2020; Ferreira and Almeida, 2021; Savla and Levy, 2020; Negrini and R. Riccardi, 2018). Table 3.2 shows some quotes from the interviewees. All interviewees in this study stated the use of rule-based transaction monitoring within their institution. The rule-based approach is based on scenario-based rules that trigger alerts when a particular event occurs. Along with the rule-based methods, a lot of statistical analysis is conducted to set thresholds for the systems which is a very complex and time-consuming task for institutions (Chau and Dijck Nemcsik, 2020).

While some banks add supplementary tools to the existing method, the majority of solutions still rely solely on the rule-based method. For instance, some institutions are incorporating entity resolution and network analysis with the help of external companies to enhance their transaction monitoring approach. Although external companies improve the current approaches there are still inefficiencies, due to the fact it's being used with rule-based methods, as mentioned by a specialist.

In contrast, an interviewee who operates as the Head of Financial Crime Detection at one of the largest banking institutions revealed that they have started testing and incorporating machine learning, moving away from rule-based monitoring. Machine learning is being adopted to detect anomalies in customer behaviour at the transactional level, allowing them to search for indications of money laundering and terrorist financing.

One of the AML specialists described the anomaly detection process by comparing it to searching for a needle in a haystack. Instead of looking for the needles in the haystack like traditional approaches, the institution defines what normal looks like and flags transactions that deviate from it. This approach utilises the vast amount of normal transaction data that the bank possesses (92 petabytes). However, when developing this approach the institution needs to be extremely cautious of including illegal transactions in the group of normal transactions, as this could lead to criminals avoiding detection. Overall, the findings align with those of (Alexandre and Balsa, 2023), indicating that while rule-based methods (some with complementary components) remain widely used, some institutions are beginning to adopt machine learning techniques for detecting money laundering and terrorist financing.

3.4.2 Challenges in Transaction Monitoring

This section identifies the problems and challenges that financial institutions encounter in transaction monitoring. Table 3.3 concentrates on challenges concerning the current transaction monitoring methods utilised in the industry. Crucial issues such as high volumes of false positives, being ineffective and slow, struggling to identify new risks, and being costly are highlighted by experts. Institutional challenges affecting transaction monitoring are presented in Table 3.4. Some challenges include, AML departments working in isolation, loss of valuable information in investigations, poor data quality and accuracy, and the difficulty of transitioning to machine learning from rule-based approaches are stated.

Challenges of the Current Methods

The findings show that one key problem in transaction monitoring that was repeatedly mentioned is the high number of false positives that are generated. It is difficult to filter out false positives from large volumes of transactions, especially when certain risks need to be captured. The industry average for false positives, which has been improving in recent years, was mentioned to be around 90-95%, in line with the current literature (Ketenci et al., 2021).

A high false positive rate leads to multiple issues in banks and is a major problem that has to be addressed to enhance detection. It negatively impacts the institution by increasing operational costs and the number of resources required to investigate alerts. As institutions need to investigate every alert the more false positives lead to more employees analysing alerts.

Another challenge that arises due to spending large amounts of resources on backlogs and false positives is the lack of ability and time to focus on new emerging risks. This will cause problems in the future as banks will not be able to adapt to criminals evolving techniques and fail to detect fraudulent transactions. The issue of false positive alerts also impacts customer experience, as these alerts often result in additional information required from customers. This can negatively impact customer satisfaction, particularly when the alert is a false positive. An additional problem that causes customer dissatisfaction is during the production of rules, events, and scenarios for the current rule-based methods, which can result in losing customers.

Identifying the effectiveness and efficiency of the current transaction monitoring methods is another challenge within financial institutions. It is crucial to understand how and what

Table 3.3: Quotes on the problems and challenges financial institutions face with current transaction monitoring methods.

	What are the problems and challenges of current transaction monitoring methods?
1	*“A major problem is there are certain risks you need to capture and it’s very hard to filter out things that are likely to be false positives”*
2	*“The volume of false positives. Industry-standard, I think, is slightly improving now, it’s about 90 to 95%”*
3	*”Every risk you pick up means an alert and investigation. This has a huge cost on the investigation side, operational cost basically”*
4	*“We are constantly trying to keep down on backlogs and false positives, so focusing on new and emerging risks doesn’t become the focal point”*
5	*”It’s very hard for companies to get a good understanding of how effective or efficient their monitoring solution is”*
6	*“The major challenge in banks and transaction monitoring systems is actually keeping up with new types of crimes”*
7	*”So currently because we’re producing many events for the rule-based system, it means we need to ask our customers many questions. So that creates a huge amount of customer dissatisfaction”*
8	*”Customer segmentation is a very clumsy, very 1990s technology because the first of these monitoring engines were built around the early 2000s, so they’re reliant on that 1990s view of the world, and it still haunts us today”*
9	*”Having to spend a lot of money on constantly refreshing your segmentation, constantly re-tuning, constantly testing that your thresholds are accurately set, checking you’re below the line testing and wide open testing to see whether you’re missing things that you should be picking up erodes the accuracy of the traditional rule-based method you’re setting up”*
10	*”It is challenging to monitor dual-use goods transactions. They are called duel goods as you can buy metal for good purposes but you can also use metal to make arms”*

your transaction monitoring model covers to develop and enhance it. Uncertainty regarding the type of risks the institution covers and lack of clarity can be very problematic and costly to institutions considering the high standards regulators set. The rule-based methods’ slow speed at processing transactions is also identified as a problem. Slow processing time can impede efficiency and effectiveness in identifying potential risks, which can result in a regulatory cost for banks. Another issue with current rule-based methods that can cause problems with regulators is the ability to keep up with new and emerging types of money laundering and terrorist financing, which is also mentioned in the literature (Dill, 2021). The inability of rule-based models to adapt to changes in customers’ behaviour is an additional weakness, as a larger number of false positives can be generated. To compound this problem, maintaining the rules and performing effective testing is complex and costly for institutions. Also, the difficulty of motioning duel use goods was found during this study as well as cross-bored transactions, which is also stated in the current literature (Collin, Cook, and Soramaki, 2017). It is challenging to distinguish between using duel-use goods for legitimate reasons or potentially dangerous ones.

The semi-structured interview findings suggest that some specialists believe that customer segmentation within the traditional rule-based monitoring methods and threshold setting is outdated and cumbersome technology that does not accurately reflect customer behaviour. Customer segmentation is seen as a 1990s type of technology and an old way of thinking that negatively impacts the transaction monitoring domain today. The complex nature of customers' behaviours, and the challenge of accurately segmenting customers, further complicates the issue. Therefore, these challenges result in a high rate of false positives, an inaccurate view of customers' behaviours, and higher costs for re-tuning and testing.

Challenges Within the Institutions

Various departments in financial institutions have significant indirect impacts on the transaction monitoring process. One major issue is that KYC information is often outdated and inaccurate, leading to either over-alerting or under-alerting on particular individuals. This negatively impacts the customer segmentation process and reduces the accuracy of the overall process. Additionally, the departments in AML (i.e. KYC) need to be embedded within the transaction monitoring and work together, not in isolation. Working in cohesion will greatly impact the accuracy and efficiency of the entire AML process. Another issue raised by a specialist is that organisations have to be cautious of the resources and expertise that are given to upstream processes, as it has a positive impact on downstream processes like transaction monitoring. The investigation process in AML also needs improvement as there is often not enough information on why an alert is closed, which limits the ability to learn from these cases by a feedback loop.

Data is a challenge in institutions and there is a need for a "golden source" of internal data that provides a single customer source with accurate and up-to-date information about the customer. The fact banks often have multiple sources of data from various departments and products adds to the single data source issue. The interviewees also highlight the impact of data quality and inaccuracy on transaction monitoring, with discrepancies between different customers' data, due to evolving regulators, creating problems for rule-based methods. Onboarding customers in the past differentiates from now and if the data is not updated regularly problems occur. Many AML experts included in this study highlighted the data quality issue impacting transaction monitoring. Incomplete data sets and data complexity present challenges to transaction monitoring, especially in the context of building new models. Many research articles in the literature have mentioned data issues in transaction monitoring (Halter et al., 2011). Finally, the interviews indicate that monitoring customers across multiple jurisdictions presents data challenges, as not all countries allow data sharing. Data privacy was also identified in the literature as a problem ((Milaj and Kaiser, 2017), as it forces institutions to set up different models in different jurisdictions.

Another challenge in institutions is understanding how to move away from the current rule-based methods and implement new solutions. The transition away from current systems will be complex and time-consuming, therefore, institutions need to bring in specialists with the necessary knowledge and skills to transition smoothly. Furthermore, the current approach in the AML departments tends to be reactive whereas banks need to be more proactive in detecting and addressing potential risks to start making a change.

Table 3.4: Quotes on the problems and challenges within the institutions affecting transaction monitoring.

	What are the problems and challenges within the institutions affecting transaction monitoring?
1	*"What we're lacking today is that KYC doesn't get updated as often, so unfortunately what happens is you over-alert or you under-alert on some people"*
2	*"KYC needs to be embedded within your transaction monitoring solution and your sanctions. However, controls are currently working in isolation but should be more aligned. So as an example for KYC, we can get them to ask questions from a transaction monitoring perspective"*
3	*"Currently you have a drop-down list of five reasons why you closed the alert, you can't really learn anything from that"*
4	*"Making sure to have a golden source for internal data, is one of the main challenges today"*
5	*"If there was an all-in-one data source then multiple data sources wouldn't be a problem"*
6	*"There's a lot of rules and scenarios which have to run on bad quality data, which is either out of date or not detailed enough"*
7	*"If you're monitoring across 58 countries like we are, not all of those countries allow data sharing between jurisdictions, so you have to set up different models in different jurisdictions. All financial institutions suffer the same"*
8	*"A big pain point is to understand how you can move from rule-based to AI. That's probably a little bit of the solution to the productivity issue too"*
9	*"A lot of the things that we're doing are reactive rather than proactive, so we're not necessarily harnessing the data and understanding where it starts to change"*

3.4.3 Requirements

This section presents the features and capabilities that are required by a new transaction monitoring method, found during this study, to produce efficient results and be successfully deployed in the industry. The requirement features are then analysed and discussed. Table 3.5 presents the quotes from the interviewees on the requirements for a new transaction monitoring method. Additionally, an importance score given by the specialists for various features, along with an average rating for each feature, is presented in Table 3.6. Each feature is described as follows:

- Explainability – the method's ability to be able to explain or interpret why it has flagged a transaction as suspicious;

- Flexibility – the method's ability to adapt to changes in customers' behaviours over long periods;

- Detection speed – how quick the method can identify the suspicious activity from when it took place;

- Scalability – the number of transactions that can be processed by the new approach;

- Customer experience – how important the customer experience is in transaction monitoring.

Explainability and Effectiveness

Explainability and effectiveness were identified as a requirement for a new transaction monitoring method by multiple specialists during the semi-structured interviews. An average importance score of 9.1 was given for explainability showing the value it has in the transaction monitoring domain. Explainability is crucial to satisfy regulatory and audit requirements is any issues arise. The importance of transparency in methods for transaction monitoring was further stressed as institutions need to explain how it works and the risks they are covering to regulators. Therefore, the explainability of the method is crucial and can be achieved by understanding what it does, why it was built, and how it works. Along with the explainability of how the method detects transactions, showing the method's effectiveness is equally as important. Overall, explaining how the method reaches its outcomes and proving effectiveness are crucial requirements for a new transaction monitoring method (Kute et al., 2021).

Flexibility

During the interviews, the specialists highlighted the value and necessity of flexibility in a new transaction monitoring method and gave an average importance score of 7.6. Adapting to unexpected changes such as covid can facilitate a more stable model capable of attaining a higher accuracy regularly. Having a flexible method will also reduce the number of false positives during seasonality or periods when customers' behaviours are changing (Bolton and D. J. Hand, 2002). Flexibility will give an institution the ability to customise the detection process and adapt to future changes in transactions and regulations.

Detection Speed

The AML specialists involved in this study gave an average importance score of 5.5 for detection speed for a new transaction monitoring method. The findings show that participants agreed that while detection speed is important, it is not always the most critical factor (Sterling, 2015). In transaction monitoring to detect a suspicious customer sometimes you need to look at the customer's transactions over a long period to identify if their patterns are suspicious. Transactions are linked and related to each other, therefore, context is required before alerting a report which can take an extended period of time reducing detection speed. The finding indicates the importance of taking the time to analyse transactions and the need for a balance between speed and accuracy for a new transaction monitoring method. While regulators value quick detection speed, they understand that transaction monitoring is a complex process. Ultimately, the main purpose is to accurately identify and stop the maximum amount of money launderers.

Scalability

The requirement of scalability for a new transaction monitoring method is evident from the quotes of the interviewees and the importance rating given by the experts. Financial institutions

Table 3.5: Quotes on the requirements for a new transaction monitoring method.

	What is required from a new transaction monitoring to be successful in the industry?
1	*”Explainability of the method is key because if there are any regulatory or audit issues, you’ve got to be able to explain your method”*
2	*”I think the main problem is that regulators are not convinced with it (artificial intelligence). So you really need to be able to know what you’re building, why you’re building it, what your objectives are, what your risks are, and how you can explain it as well”*
3	*“If the method can’t adapt properly you will get a whole bunch of false positives due to seasonality or changes of behaviour”*
4	*”Transactions are only relevant in the context of other transactions, so you might have scenarios that work based on looking at three months’ worth of transactions and then seeing how they relate to each other. Detection speed is important, but you may not always pick up your scenarios. A transaction tool may see three transactions that are similar and then the third aggregates all three and reports that”*
5	*”Banks process millions of transactions every day. So when running a detection capability across a large bank everything has to be scalable”*
6	*“As we are producing a lot of events for the rule-based system, it means we need to ask our customers a lot of questions. So that creates a huge amount of customer dissatisfaction”*
7	*”When the transaction gets flagged and goes into the investigation stage, we learn a lot of information, you can reuse that information intelligence to improve detection”*
8	*“I’m a part of the X regulatory group on monitoring, and in the recent meeting we had, it was clear that institutions are still struggling with the accuracy of monitoring”*
9	*”I think it’s going to be quite critical to identify any new risk that we can’t think of. Can we get machines to identify unusual behaviours in the data compared to its peers or a group of people, and show new risk”*
10	*”To be able to successfully implement a new method you need to consider the resources required, how can an effective team size manage it? Implementation experience of that team?”*

Question: On a scale of 1-10, how important are the features listed below for a transaction monitoring method in AML and why?

Features	Specialist								Average Score
	1	2	3	4	5	6	7	8	
Explainability	9	9	8	10	10	8	9	10	9.1
Flexibility	8	10	7	7	9	5	8	8	7.6
Detection Speed	7	5	5	5	6	2	8	7	5.5
Scalability	10	9	8	8	9	5	7	10	8.2
Customer experience	9	1	4	4	1	1	9	1	2.6

Table 3.6: Importance of features given by specialists for transaction monitoring methods in AML

are required to process millions of transactions daily. The findings demonstrate that scalability is crucial for a new transaction monitoring method to handle the amount of data produced daily. Additionally, an average important score of 8.2, reinforces the significance and need for scalability. A scalable method will make it possible to deploy the approach in the industry and enable the identification of suspicious activities in large datasets (Han, Y. Huang, S. Liu, et al., 2020).

Customer Experience

Mixed responses were given on how transaction monitoring affects the customer experience. This was due to various institutions having different approaches and processes for transaction monitoring. A majority of the specialists did not regard the customer experience as very crucial, as the AML process did not impact the customers. The customer only gets affected post transaction monitoring, specifically during the investigational stage. However, some institutions contact customers and ask questions during or before the transaction monitoring process to create accurate events and rules, which can cause dissatisfaction. An average importance rating of 2.6 was given for customer experience with a large range due to the differences in financial institutions.

Feedback Loop

AML specialists stressed the importance of reusing investigation outputs through a feedback loop as it's the best source of additional data. This finding highlights the potential use of data obtained during the investigation stage to continually improve the detection accuracy in transaction monitoring methods (Khritankov, 2021). This approach can reduce false positives drastically and provide individual customer-level monitoring that can adapt to specific customers' behaviours. Financial institutions hold a lot of intelligence that should be used to enhance transaction monitoring further and more research is required in this field.

Higher Accuracy

There is a need for higher accuracy in current transaction monitoring approaches, due to the high occurrence of false positives and a lack of efficiency (J.-J. Rocha-Salazar, M.-J. Segovia-Vargas, and M.-d.-M. Camacho-Miñano, 2021). The requirement of attaining a higher accuracy for a new approach was crucial for the specialists. The findings suggest that the AML sector is seeking to attain greater accuracy for detection. An interviewee in the financial sector and a regulatory group stated emphasised the challenges around detection in the AML field. These results highlight the limitations of traditional methods and the need for advanced machine learning models.

Identification of New Risks

The AML specialists highlighted the importance of identifying any new risks that can emerge in transactions. Along with identifying new risks, the importance of identifying hidden relationships was mentioned. Meeting this requirement would not only improve the productivity of transaction monitoring methods but also ensure compliance with regulatory requirements, preventing potential fines, and damage to reputation (Pellegrina and Masciandaro, 2009). However, keeping up with the ever-changing landscape of criminals' complex money laundering techniques is very challenging and costly.

Efficient Customer Segmentation

An efficient customer segmentation process is thought of as a requirement for a new transaction monitoring method by some specialists included in this study. An efficient and accurate customer segmentation process can positively impact the accuracy of detecting money launderers in multiple ways. Customers can be monitored based on tailored risk profiles and their different behavioural patterns. Despite specialists' scepticism towards clustering customers, they recognised the significance of customer segmentation if done right.

Implementation

The understanding of implementation was identified as a requirement for a new transaction monitoring method. The findings highlighted the importance of considering how a new method can fit and be implemented into an institution and recognise the numerous challenges that can arise during the integration process into existing systems and processes (Zolkaflil, Omar, and Syed Mustapha Nazri, 2019). Although developing an effective transaction monitoring method is crucial, having a well-equipped team with the right skills and experience to manage the implementation process is equally important.

Handling Seasonality

During the semi-structured interviews, the specialists emphasised the need for a new method to account for seasonality. During months or times when unexpected activity is going to take place, it should be able to handle it, learn from previous situations, and not produce a huge amount of false positives. However, the new method must still be able to detect suspicious behaviour within the context of expected patterns. Overall, to enhance efficiency and reduce the occurrence of false positive alerts the new method should be able to adjust to seasonality while identifying suspicious activity.

3.4.4 Artificial Intelligence and Machine Learning for Transaction Monitoring

When AML specialists were asked about the role of artificial intelligence and machine learning in the field of transaction monitoring it was clear that the interviewees believed the domain was heading towards adopting these technologies. The findings of this research are in line with the current literature, as proven by the number of articles released on machine learning for transaction monitoring in recent years (Kute et al., 2021). Table 3.7 shows the quotes from the interviewees about their opinions on using artificial intelligence and machine learning for transaction monitoring.

The specialists stated that artificial intelligence is the way forward and will make processes under transaction monitoring easier to manage. Machine learning techniques can better handle

Table 3.7: Quotes from the interviewees on their opinions on AI and machine learning for transaction monitoring.

	What are your opinions on artificial intelligence and machine learning for transaction monitoring?
1	*"I think artificial intelligence and machine learning is the only way forward for us"*
2	*"AI is now an absolutely key part. If you get your AI right, then everything else under that will be far easier to manage"*
3	*"I think at the moment we're in a transition stage from very rigid, strict, mature rule-based systems and now moving towards a more analytical platform with advanced capabilities such as entity resolution, machine learning, and graph analytics"*
4	*"Problems relate a lot to the rule-based solutions versus AI and I think banks are behind with AI generally"*
5	*"I think it needs to be a bit clearer on what the industry means by AI and machine learning and how it can be applied because it has been thrown around for quite some time without much change in the industry. How can AI be used and harnessed? I'm a bit sceptical of it"*
6	*"I think being able to tell, just with the transaction data (using machine learning) is a long way off for the industry"*

the complexities of detecting money launderers and can update systems quicker and more efficiently than current rule-based methods (Z. Chen, Van Khoa, Teoh, et al., 2018). This study also implies that a transition from very strict and rigid rule-based systems to more advanced capabilities such as entity resolution, machine learning, and graph analytics is taking place in transaction monitoring. Although a transition towards artificial intelligence and machine learning is happening, banks are still behind in adopting such technologies. Issues such as not having the necessary infrastructure and shortage of skilled and knowledgeable staff in the machine learning field to facilitate and manage systems prevent financial institutions from implementing artificial intelligent lead solutions. Overall, the specialists are optimistic about machine learning solutions, but further work is needed to fully leverage the potential of these technologies.

One specialist expressed some scepticism towards the implementation and use of machine learning in the transaction monitoring domain. The expert highlighted the need for greater clarity of artificial intelligence implications for the industry as it has been a discussed topic for an extended period without substantial change. They believed that detecting money launderers just through transactional data is "a long way off" for many banks. Although this particular interviewee gave diverging viewpoints compared to the other expert's opinions they recognised the relevance and potential of machine learning in this domain.

3.4.5 Future of the Domain

During the interviews, AML specialists discussed their perspectives on the future of the transaction monitoring domain. Table 3.8 presents thematic analysis results on the future of the transaction monitoring domain. It was believed that in the future regulators may expect en-

Table 3.8: Quotes of interviewees on the future of the transaction monitoring domain.

	Where do you see the future of the transaction monitoring domain heading?
1	*"In three years' time, regulators can say you had the technology, data, and the people with the capability and skills to develop all of this, but you didn't build an artificial intelligence or machine learning lead method. Why not?"*
2	*"In the past, regulators were a bit nervous about automated transaction monitoring which is currently used, but within five years the same regulators were telling its major banks they had to use automated transaction monitoring capabilities for customer monitoring. If they weren't then they needed to be able to explain why. I think in five years' time we'll have the same regulators, who are nervous about machine learning capabilities now but will be expecting it in the future"*
3	*"The expectation will be that, we don't retrospectively look at what's happened, but we use the technology we have to look forward and say we think this provides us with a risk that we're not willing to take, even though we don't have any evidence to suggest that it is a crime. However, you have evidence to suggest that it could be in the future. Now that's where the technology is taking us, whether we like it or not"*
4	*"A regulator looking backward will say, you developed machine learning capabilities that were capable of predictive behaviour and predictive modelling. You had the capability of the people and the analysts in your bank to do that and you didn't use it to predict that this particular subset of customers were terrorists. Then you have done something bad and it can have problems with the regulators"*

hanced transaction monitoring capabilities and can question institutions that have not adopted or incorporated them.

Institutions with access to technology, data, and skilled personnel will be required to develop artificial intelligence or machine learning solutions. Another interviewee added that regulators will expect the adoption of machine learning in the future by explaining that historically regulators once hesitated about automated transaction monitoring, yet in five years, the same regulators began mandating its use by major banks (Dill, 2021). This insight implies that machine learning is likely to be used for transaction monitoring in the future, which agrees with the existing literature, although regulators are currently sceptical. Therefore, it may be crucial for financial institutions to adopt machine learning as early as possible to prevent potential problems down the line. In the future, once machine learning and artificial intelligence are adopted by financial institutions, one specialist believes that transaction monitoring will be used to predict criminals and prevent their actions before the crime takes place. Overall, it is thought that machine learning will be a key component in transaction monitoring and evolve the domain further.

3.4.6 Transaction Monitoring Approaches

This section outlines and explores the transaction monitoring method identified by AML specialists during the interviews. It discusses different approaches that can improve the current

Table 3.9: Quotes of interviewees on the type of approach for producing new transaction monitoring methods.

	What type of approach can be used when producing a new transaction monitoring method?
1	*“There are smaller components that could be enhanced by machine learning first and foremost. The first way could be alert prioritisation, understanding where the biggest risks lie so that more attention and focus can be prioritised on that area”*
2	*”Using machine learning to hibernate alerts to detect and tell us you’re probably going to close this alert without raising a SAR so we shouldn’t actually investigate it any further”*
3	*”Currently a lot of tricky rules are trying to identify hidden relationships. So how could you use entity resolution and work out what the key links between different parties or different sorts of companies are, to allow you to do advanced hidden relationship identification? That leads to graph analytics. So focusing on patterns rather than transactions, to find hidden networks and relationships”*
4	*”Look at what is normal and then at the things that are outside of that? Next use outlier analysis and aggregation to create a risk view of a particular customer and then based on that aggregation of risk, you can decide whether to look at them through an investigation or not”*
5	*”With scorecards, you get interested in a transaction because of risk. Then you look at other factors and risks of the transaction, giving it a score, and if it adds up to a certain score you decide to flag it”*

transaction monitoring methods. One specialist suggested utilising machine learning as an add-on tool for current rule-based methods to hibernate alerts. Other methods such as anomaly detection and graph machine learning were also suggested. Additionally, a scorecard approach to prevent money laundering was proposed. Table 3.9 presents the key quotes from the interviews.

An approach suggested by multiple specialists is to incorporate machine learning as an add-on tool to the current rule-based methods. The experts offered valuable insights into the potential use of machine learning as a hibernation or alert prioritisation model. Alert prioritisation can provide institutions with a deeper understanding of high-risk areas allowing for more attention and resources to be allocated accordingly. Additionally, it can be an easy way to initiate machine learning adoption in institutions, while being a cost-effective and easily implementable approach. A risk migrating and alert hibernation approach was also proposed to reduce the number of false positives and the number of redundant investigations. Utilising historical suspicious activity reports and past experiences with unsuccessful alerts, machine learning can recommend, with a degree of confidence, whether to investigate a specific transaction. In the current literature, several researchers have proposed alert prioritisation and hibernation methods for transaction monitoring (Astrova, 2023). Although these approaches benefit institutions in multiple ways, they may not be sustainable in the long term, as they build upon an existing, potentially inefficient method.

Graph analysis approach was identified by the specialists as a potential transaction monitoring method. An interviewee's quotes highlight the need to move away from traditional rule-based detection and towards a more advanced approach that leverages graph analytics to identify hidden relationships. It is believed graph analysis should focus on patterns instead of individual transactions to identify hidden networks and relationships. Graph machine learning constitutes a scalable methodology capable of handling large volumes of data while conducting holistic analysis which can address the high false positive issues creating an efficient approach. Overall, the specialists acknowledge the value of utilising graph analysis for monitoring but understand the difficulty of successfully achieving such an approach. Challenges such as the complex routing of money through entities and other financial markets by criminals make it difficult to identify hidden relationships. While the application of graph machine learning has been explored in the current literature, further research is required for advancements (Kute et al., 2021).

AML specialists identified anomaly detection using machine learning as another approach to enhance transaction monitoring. Detecting money launderers can be improved by training a machine learning model on a dataset composed of "normal transactions" and then flagging transactions that deviate from those normal transactions as suspicious. Given the vast amount of transactions that are deemed "normal" in financial institutions, an anomaly detection approach can yield promising results, however, it can be computationally expensive. An anomaly detection and data-driven method is thought to be able to increase the accuracy and reduce the amount of false positives compared to the existing methods in the industry. The "normal transactions" dataset must be large and of high quality to attain desirable and accurate results, otherwise, fraudulent transactions can process undetected. An autoencoder technique (Zhiyuan et al., 2021) is published in the literature with a similar methodology along with many other anomaly detection approaches for transaction monitoring (Z. Chen, Van Khoa, Teoh, et al., 2018).

During the semi-structured interviews specialists stated that they were exploring and working on a scorecard model for transaction monitoring. They believed that a scorecard model would allow for a more holistic monitoring of transactions and consider multiple factors and risks of the transactions before making a decision. A specific set of rules or scenarios with different weights based on risk is required to build a scorecard approach. For each rule, a transaction triggers an individual score is given. Subsequently, these scores are aggregated to determine if the transaction should be alerted. A benefit of the scorecard method is that it allows the institution to control and reduce the volume of false positives produced. This method could improve current transaction monitoring methods, however, it would be very difficult to set accurate weights for the rules, which could lead to low accuracy. The scorecard approach could be explored further as there is limited research in the existing literature due to most efforts focusing on machine learning solutions.

3.5 Summary

In conclusion, this study explored the current transaction monitoring approaches employed in the industry, as well as the challenges associated with them. Furthermore, the research delved into artificial intelligence and machine learning in transaction monitoring, analysing the specialists' opinions and the future of the domain. In addition, requirements for new solutions are presented along with specialists' opinions on potential future transaction monitoring approaches.

The findings of this research suggest that the current transaction monitoring methods are inefficient and highlighted the problems it brings to institutions such as high false positive alerts, low detection accuracy, and increased operational costs. A transition from strict, rule-based methods to more advanced machine learning capabilities is seen to be taking place and believed to be the solution for the future of the domain. Although there is optimism surrounding advanced capabilities, it is crucial to acknowledge that additional work and research are required to adopt these technologies. Analysing the semi-structured interviews provided requirements for a new transaction monitoring method to be successful and efficient, emphasising the need for scalability, high accuracy, proving effectiveness, regulatory compliance, and explainability. Desirable features such as information feedback loops, flexibility in customers' behaviours, and the ability to identify hidden relationships were also identified. Various approaches to enhance transaction monitoring suggested by the specialists were also explored, including add-on tools to current rule-based methods, graph analysis, and anomaly detection techniques. Additionally, a scorecard method was identified and discussed during the interviews.

This study contributes to the existing literature on transaction monitoring and anti-money laundering by conducting semi-structured interviews with specialists in the domain. Excessive analysis of experts' opinions and knowledge provides a deeper insight into the current problems and requirements for new solutions in transaction monitoring. Currently, in the literature there is a disconnect and lack of studies between industries and academicians around transaction monitoring, this study bridges the gap to gather a better understanding of what is expected and required. These findings can assist researchers and stakeholders to get an in-depth understanding of the problems in the transaction monitoring domain and provide a guideline to produce successful and efficient transaction monitoring methods to meet the industry's requirements.

In summary, improvements are required in the transaction monitoring domain to detect money laundering activities with higher accuracy and efficiency. It is crucial for researchers to continue investigating solutions with advanced capabilities. Although there are challenges to overcome and further research is needed, the AML specialists' perspectives and opinions suggest that technologies such as machine learning will greatly enhance the field. Adopting artificial intelligence and machine learning will not only improve the detection of money launderers but also pave the way for more innovative and enhanced approaches that can address the ever-changing challenges in the anti-money laundering industry.

Chapter 3 References

Alexandre, Claudio Reginaldo and João Balsa (2023). "Incorporating machine learning and a risk-based strategy in an anti-money laundering multiagent system". In: *Expert Systems with Applications* 217, p. 119500. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2023.119500.

Ali, Ahmed Hussein (2019). "A Survey on Vertical and Horizontal Scaling Platforms for Big Data Analytics". In: *International Journal of Integrated Engineering* 11.6, pp. 138–150.

Arora, C., M. Sabetzadeh, and L.C. Briand (2019). “An empirical study on the potential usefulness of domain models for completeness checking of requirements”. In: *Empirical Software Engineering* 24, pp. 2509–2539. DOI: 10.1007/s10664-019-09693-x.

Astrova, Irina (2023). “Anti-money Laundering Powered by Graph Machine Learning: ”Show Me Your Friends and I Will Tell You Who You Are””. In: *Intelligent Decision Technologies*, pp. 1–19. DOI: 10.3233/IDT-220193.

Bolton, Richard J. and David J. Hand (2002). “Statistical Fraud Detection: A Review”. In: *Statistical Science* 17.3, pp. 235–255. DOI: 10.1214/ss/1042727940.

Canhoto, Ana Isabel (2021). “Leveraging machine learning in the global fight against money laundering and terrorism financing: An affordances perspective”. In: *Journal of Business Research* 131, pp. 441–452. ISSN: 0148-2963. DOI: 10.1016/j.jbusres.2020.10.012.

Chau, Derek and Maarten van Dijck Nemcsik (2020). *Anti-Money Laundering Transaction Monitoring Systems Implementation: Finding Anomalies*. John Wiley & Sons.

Chen, Z., L. D. Van Khoa, E. N. Teoh, et al. (2018). “Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review”. In: *Knowledge and Information Systems* 57, pp. 245–285. DOI: 10.1007/s10115-017-1144-z.

Collin, Matthew, Samantha Cook, and Kimmo Soramaki (2017). *The Impact of Anti-Money Laundering Regulation on Payment Flows: Evidence from SWIFT Data*. Tech. rep. 445. Center for Global Development, p. 52.

Creswell, John W. and J. David Creswell (2018). *Research Design*. 5th ed. SAGE Publications.

Dehouck, Maja and Marieke de Goede (2021). *Public-Private Financial Information-Sharing Partnerships in the Fight Against Terrorism Financing*. University of Amsterdam. 6-30.

Dicicco-Bloom, B. and B. F. Crabtree (2006). “The qualitative research interview”. In: *Medical Education* 40, pp. 314–321. DOI: 10.1111/j.1365-2929.2006.02418.x.

Dill, A. (2021). *Anti-money laundering regulation and compliance: Key Problems and Practice Areas*. Cheltenham, UK: Edward Elgar Publishing Limited.

Eifrem, Emil (2019). “How graph technology can map patterns to mitigate money-laundering risk”. In: *Computer Fraud & Security* 2019.10, pp. 6–8. DOI: 10.1016/S1361-3723(19)30105-8.

FATF (2023). *International Standards on Combating Money Laundering and the Financing of Terrorism & Proliferation*. Financial Action Task Force. Available from: http://www.fatf-gafi.org/recommendations.html [Accessed 12 May. 2023].

Ferreira, L. and C. Almeida (2021). “AML Transaction Monitoring Systems: An Overview”. In: *Journal of Financial Crime* 28.2, pp. 355–372.

Halter, Emily Marie et al. (2011). *The puppet masters: how the corrupt use legal structures to hide stolen assets and what to do about it*. Tech. rep. Washington, D.C.: World Bank Group.

Han, J., Y. Huang, S. Liu, et al. (2020). “Artificial intelligence for anti-money laundering: a review and extension”. In: *Digital Finance* 2, pp. 211–239. DOI: 10.1007/s42521-020-00023-1.

Hasan, M. M., J. Popp, and J. Oláh (2020). “Current landscape and influence of big data on finance”. In: *Journal of Big Data* 7, p. 21. DOI: 10.1186/s40537-020-00291-z.

Johnson, M. C. and S. R. Kessler (2019). “The art and science of semi-structured interviewing: A comprehensive guide for researchers”. In: *Qualitative Research Journal* 21, pp. 131–147. DOI: 10.1177/1468794119825569.

Ketenci, Utku Gorkem et al. (2021). "A Time-Frequency Based Suspicious Activity Detection for Anti-Money Laundering". In: *IEEE Access* 9, pp. 59957–59967. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3072114.

Khritankov, A. (2021). "Hidden Feedback Loops in Machine Learning Systems: A Simulation Model and Preliminary Results". In: *Software Quality: Future Perspectives on Software Engineering Quality*. Vol. 404. Lecture Notes in Business Information Processing. Cham: Springer. DOI: 10.1007/978-3-030-65854-0_5.

Kumari, P. and A. Gupta (2020). "An empirical study on the current status of anti-money laundering compliance and the role of artificial intelligence". In: *Journal of Financial Crime* 27.3, pp. 895–914.

Kute, Dattatray Vishnu et al. (2021). "Deep Learning and Explainable Artificial Intelligence Techniques Applied for Detecting Money Laundering–A Critical Review". In: *IEEE Access* 9, pp. 82300–82317. DOI: 10.1109/ACCESS.2021.3086230.

Liu, Xuan and Pengzhu Zhang (2008). "Research on Constraints in Anti-Money Laundering (AML) Business Process in China Based on Theory of Constraints". In: *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*, pp. 213–213. DOI: 10.1109/HICSS.2008.374.

Mekpor, E.S. (2019). "Anti-money laundering and combating the financing of terrorism compliance: Are FATF member states just scratching the surface?" In: *Journal of Money Laundering Control* 22.3, pp. 451–471. DOI: 10.1108/JMLC-09-2018-0057.

Milaj, Jonida and Carolin Kaiser (Apr. 2017). "Retention of data in the new Anti-money Laundering Directive—'need to know' versus 'nice to know'". In: *International Data Privacy Law* 7.2, pp. 115–125. ISSN: 2044-3994. DOI: 10.1093/idpl/ipx002.

Negrini, M. and R. Riccardi (2018). "Rule-Based Transaction Monitoring for Anti-Money Laundering: A Review of the Current State and Future Perspectives". In: *Journal of Financial Crime* 25, pp. 417–432.

Oztas, B. et al. (2023b). "Perspectives from Experts on Developing Transaction Monitoring Methods for Anti-Money Laundering". In: *2023 IEEE International Conference on e-Business Engineering (ICEBE)*. IEEE. Sydney, Australia, pp. 39–46. DOI: 10.1109/ICEBE59045.2023.00024.

Oztas, Berkan et al. (2022). "Enhancing Transaction Monitoring Controls to Detect Money Laundering Using Machine Learning". In: *2022 IEEE International Conference on e-Business Engineering (ICEBE)*, pp. 26–28. DOI: 10.1109/ICEBE55470.2022.00014.

Palinkas, L. A. et al. (2015). "Purposeful sampling for qualitative data collection and analysis in mixed-method implementation research". In: *Administration and Policy in Mental Health and Mental Health Services Research* 42, pp. 533–544. DOI: 10.1007/s10488-013-0528-y.

Pellegrina, Lucia Dalla and Donato Masciandaro (2009). "The Risk-Based Approach in the New European Anti-Money Laundering Legislation: A Law and Economics View". In: *Review of Law & Economics* 5.2, pp. 931–952. DOI: 10.2202/1555-5879.1422.

Pontes, R. et al. (2022). "Anti-money laundering in the United Kingdom: new directions for a more effective regime". In: *Journal of Money Laundering Control* 25.2, pp. 401–413. DOI: 10.1108/JMLC-04-2021-0041.

Rocha-Salazar, José-de-Jesús, María-Jesús Segovia-Vargas, and María-del-Mar Camacho-Miñano (2021). "Money laundering and terrorism financing detection using neural networks and an abnormality indicator". In: *Expert Systems with Applications* 169, p. 114470. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2020.114470.

Savla, R. and J. Levy (2020). "A Critical Review of AML Transaction Monitoring: Challenges and Opportunities". In: *Journal of Money Laundering Control* 23, pp. 447–463.

Simonova, A. (2011). "The risk-based approach to anti-money laundering: problems and solutions". In: *Journal of Money Laundering Control* 14.4, pp. 346–358. DOI: 10.1108/13685201111173820.

Sterling, S. (2015). "Identifying money laundering: Analyzing suspect financial conduct against the speed, cost, and security of legitimate transactions". In: *Journal of Money Laundering Control* 18.3, pp. 266–292. DOI: 10.1108/JMLC-08-2014-0025.

Stevens, E. A. (2020). "Understanding the inductive approach: Benefits and applications in research". In: *Journal of Research Methods and Applications* 16, pp. 1–18. DOI: 10.1080/15546128.2020.1758346.

Tertychnyi, P. et al. (2020). "Scalable and Imbalance-Resistant Machine Learning Models for Anti-money Laundering: A Two-Layered Approach". In: *Enterprise Applications, Markets and Services in the Finance Industry. FinanceCom 2020. Lecture Notes in Business Information Processing.* Vol. 401. Springer. DOI: 10.1007/978-3-030-64466-6_3.

Thomas, D. R. (2006). "A general inductive approach for analyzing qualitative evaluation data". In: *American Journal of Evaluation* 27, pp. 237–246. DOI: 10.1177/1098214005283748.

Veyder, F. (2003). "Case study: Where is the risk in transaction monitoring?" In: *Journal of Financial Regulation and Compliance* 11.4, pp. 323–328. DOI: 10.1108/13581980310810606.

Viritha, B., V. Mariappan, and V. Venkatachalapathy (2015). "Combating money laundering by the banks in India: compliance and challenges". In: *Journal of Investment Compliance* 16.4, pp. 78–95. DOI: 10.1108/JOIC-07-2015-0044.

Wronka, C. (2022). ""Cyber-laundering": the change of money laundering in the digital age". In: *Journal of Money Laundering Control* 25.2, pp. 330–344. DOI: 10.1108/JMLC-04-2021-0035.

Zavoli, Ilaria and Colin King (2021). "The Challenges of Implementing Anti-Money Laundering Regulation: An Empirical Analysis". In: *The Modern Law Review* 84.4, pp. 740–771. DOI: 10.1111/1468-2230.12628.

Zhiyuan, Chen et al. (2021). "Variational Autoencoders and Wasserstein Generative Adversarial Networks for Improving the Anti-Money Laundering Process". In: *IEEE Access* 9, pp. 83762–83785. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3086359.

Zolkaflil, S., N. Omar, and S. N. F. Syed Mustapha Nazri (2019). "Implementation evaluation: a future direction in money laundering investigation". In: *Journal of Money Laundering Control* 22.2, pp. 318–326. DOI: 10.1108/JMLC-03-2018-0024.

Chapter 4

Anti Money Laundering Transaction Dataset Generation: SAML-D

Version of record available at: B. Oztas, D. Cetinkaya, F. Adedoyin, M. Budka, H. Dogan, and G. Aksu. "Enhancing Anti-Money Laundering: Development of a Synthetic Transaction Monitoring Dataset," in 2023 IEEE International Conference on e-Business Engineering (ICEBE), Sydney, Australia, 2023, pp. 47-54.
DOI: 10.1109/ICEBE59045.2023.00028

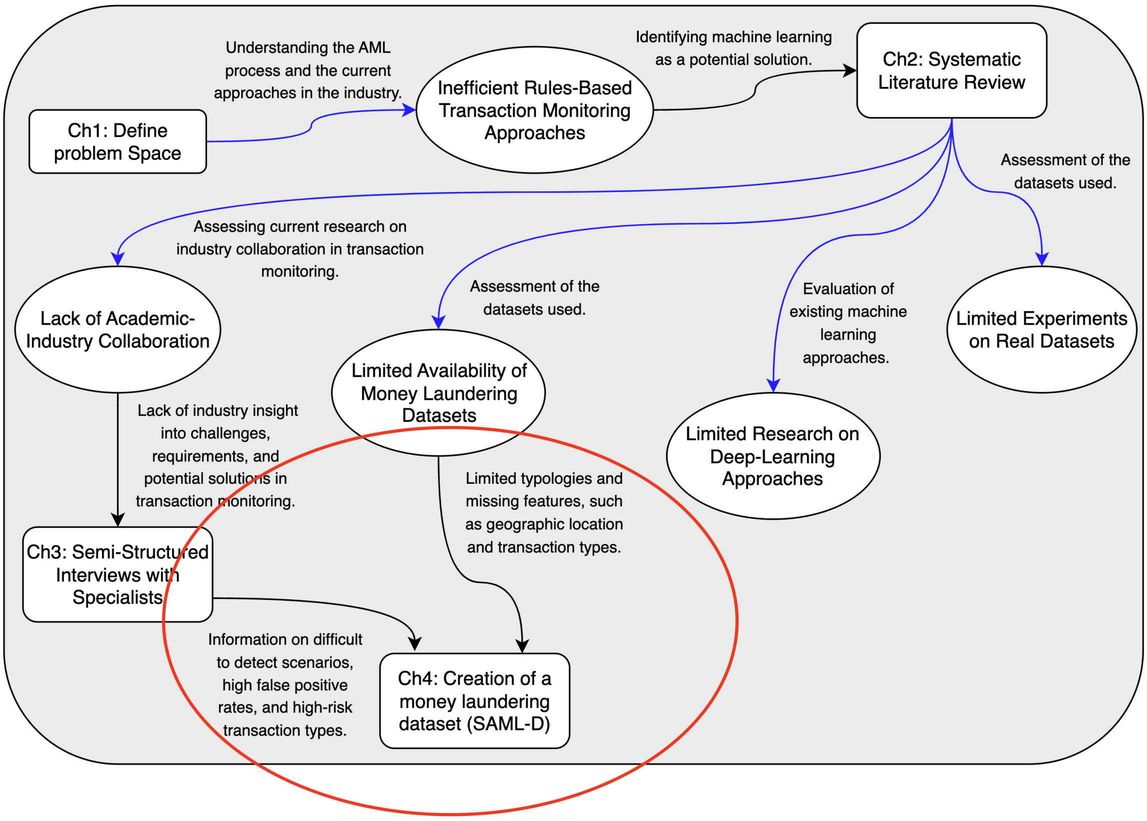

Ch1: Define problem Space
Understanding the AML process and the current approaches in the industry.
Inefficient Rules-Based Transaction Monitoring Approaches
Identifying machine learning as a potential solution.
Ch2: Systematic Literature Review
Assessment of the datasets used.
Assessing current research on industry collaboration in transaction monitoring.
Lack of Academic-Industry Collaboration
Assessment of the datasets used.
Evaluation of existing machine learning approaches.
Limited Experiments on Real Datasets
Limited Availability of Money Laundering Datasets
Limited Research on Deep-Learning Approaches
Lack of industry insight into challenges, requirements, and potential solutions in transaction monitoring.
Limited typologies and missing features, such as geographic location and transaction types.
Ch3: Semi-Structured Interviews with Specialists
Information on difficult to detect scenarios, high false positive rates, and high-risk transaction types.
Ch4: Creation of a money laundering dataset (SAML-D)

Figure 4.1: Overview of the PhD (extracted from Figure 1.3)

4.1 Introduction

This chapter introduces a new AML transaction generator to address the limited availability of money laundering datasets, fulfilling objective (iii). Figure 4.1 highlights the key limitations identified in Chapter 2, along with insights gathered from the semi-structured interviews that informed the development of Chapter 4. This chapter has been published in the *2023 IEEE International Conference on e-Business Engineering* as part of the integrated thesis.

Although, machine learning approaches are showing promising results, accessing data is challenging. Real financial transaction data consisting of money laundering behaviours is not generally available, due to legal and privacy reasons (Jullum et al., 2020)(Cheng, 2021). In cases where anonymised real data is available, the labelling of money laundering transactions is limited due to a lack of ground truths, and laundering transactions often go undetected (Europol, 2017). The availability of synthetic money laundering data is limited with these datasets often having significant shortcomings, such as the absence of critical features or a lack of diverse money laundering typologies.

The dataset introduced in this chapter is in tabular format and includes various features of 'suspicious' and 'normal' transactions conducted by different entities. The key contributions of this study involve an extension of both normal and suspicious typologies in the dataset, improving upon existing synthetic alternatives. Including new features such as geographic locations and high-risk countries adds a layer of complexity and brings a greater degree of realism to the dataset. A current challenge in the domain is comparing the results of different studies and machine learning algorithms as experiments are conducted using different datasets (Berkan Oztas et al., 2022). This dataset can address these challenges by serving as a benchmark dataset for researchers, enabling comparison and consequently supporting more meaningful analysis. Additionally, I conducted a set of preliminary experiments using machine learning approaches, demonstrating the utility of the SAML-D data and establishing a comparison point.

4.2 Related Work

AMLSim project introduces a two-component method for generating synthetic banking transaction data using a multi-agent-based simulator (Suzumura and Kanezashi, 2021). The 'Transaction Graph Generator' creates accounts and attributes based on an input CSV account file and establishes basic interactions between them. Suspicious transactions are then added using another CSV alert parameter file. The 'Transaction Simulator' uses a multi-agent program to simulate transactions within the network. The simulator mimics real-world transactions, aiming for an accurate portrayal of real-life financial behaviours. Moreover, the number of transactions can be modified during generation. Several datasets with varying amounts of transactions are provided. In this chapter, I focus on the analysis of the '100Kvertices-10Medges' dataset. To the best of my knowledge, AMLSim is the only synthetic dataset utilised to asses new transaction monitoring approaches in the current literature (Tundis, Nemalikanti, and Mühlhäuser, 2021)(Amr Ehab Muhammed Shokry, Mohammed Abo Rizka, and Nevine Makram Labib, 2020). However, there are limitations to AMLSim such as its singular transaction type.

The IT-AML dataset (Altman et al., 2023) is a synthetic representation of financial transactions created via a multi-agent virtual world model. The dataset is not derived from anonymized real-world individuals, rather it represents an entirely synthetic construct where various entities interact in ways that mimic real-world financial transactions. The dataset includes a collection of good and bad actors, with the latter attempting to launder money following one of the eight incorporated typologies: Fan-Out, Fan-In, Cycle, Bipartite, Stack, Random, Scatter Gather, and Gather Scatter. The typologies from AMLSim are used. The IT-AML dataset expands on the AMLSim dataset by incorporating more transaction types and typologies, making it a valuable AML dataset. A selection of datasets is provided, with varying numbers of transactions and illicit activities included. In this study I utilised the HI-Small_Trans.csv set.

Money Laundering Data Production (MLDP) is a dataset created during a Master's project that simulates financial transactions (Mahootiha, 2020), incorporating suspicious transactions based on traditional stages of money laundering. Despite its attention to the key money laundering stages, MLDP includes a limited number of transactions and lacks detail about the different types of money laundering methods, providing room for improvement in future synthetic datasets.

Other transactional datasets have been created for the purpose of fraud detection. Although these datasets were not designed for AML purposes, they offer valuable insights such as the size and features of the datasets, due to the domain's similarities. PaySim (Lopez-Rojas and Axelsson, 2012) takes a different approach by utilising a simulator to generate synthetic financial datasets based on a month-long sample of real mobile money transactions. The method focuses on replicating the statistical patterns observed in the original dataset. The resulting dataset is vast, containing millions of transactions and unique sender/receiver IDs. The Credit Card Fraud Detection dataset (Le Borgne et al., 2022) is a real dataset from 2013 and focuses on credit card transactions made by European cardholders. The data consists of a low percentage of fraudulent transactions and provides a large number of features for each transaction. However, 28 out of the 31 features are PCA-transformed because of confidentiality issues (V1–V28), while the others are Time, Amount, and Class.

4.3 Methodology

The approach used in creating the synthetic anti-money laundering dataset was chosen to ensure well-represented money laundering typologies, interactions between agents (bank accounts), and realistic banking transactions. This section starts by explaining the identification, selection, and development process of the normal and suspicious typologies. Then the production of the selected "normal" and "suspicious" typologies are described, involving two methods as shown in Fig 4.2; the agent-based approach (S. H. Chen and Venkatachalam, 2017) and the typology-based approach (Valbuena, Verburg, and Bregt, 2008).

Identifying and selecting the various normal and suspicious typologies was a complex procedure involving multiple phases. The first phase consisted of an in-depth evaluation of the academic literature (P. He, 2010)(Irwin, Choo, and L. Liu, 2012)(Plaksiy, Nikiforov, and Miloslavskaya, 2018), the AML domain (UK Government, 2020)(Financial Action Task Force (FATF), 2023), and existing datasets (Suzumura and Kanezashi, 2021) to identify common and emerging typologies that could be included in the dataset created in this research. Further evaluation was conducted through semi-structured interviews with eight AML experts. The interviews lasted approximately 60 minutes and were conducted online. Once all the interviews were complete, the data was transcribed and anonymised for data analysis. Semi-structured interviews were selected as they allow for an in-depth understanding of participants' experiences, opinions, and knowledge related to the AML subject (Johnson and Kessler, 2019). This helped derive a better understanding of several money laundering typologies, such as: difficult-to-detect typologies using rule-based methods, typologies leading to high false positives using rule-based methods, and high-risk payment types. Based on these evaluations, many typologies were chosen to be included in the dataset (more details in Section 4.4.1). A combination of both simple and complex typologies was generated to incorporate the various levels of strategies used in the real world. During the creation of the suspicious and normal typologies, all were tested individually to ensure the desired output was being generated.

Each typology was modelled as a graph structure, where nodes represented bank accounts and edges represented transactions between them. The nodes contained bank account attributes such as the bank id, and bank location details, while the edges captured transaction-related features, including the transaction amount, currency used, payment type, timestamp, and more. Since each typology represents a different type of financial behaviour, the number of nodes and edges varied depending on the specific typology. Additionally, the transaction attributes (edges) differed across typologies, incorporating variations in transaction date distributions, currency usage, frequency of transactions, transaction amounts, and transaction types. More details can be found in section 4.4.1. Each graph structure was responsible for generating transactions that adhered to a specific typology for a given bank account number. This ensured that the generated dataset accurately captured normal and money laundering behaviours while maintaining structural diversity across different typologies.

To generate realistic banking data, random 10-digit bank account numbers were first created to serve as unique identifiers for each account. Once the accounts were generated, each was randomly assigned a set of normal typologies, ensuring that different accounts exhibited diverse financial behaviours. These typologies represented real-world banking activities such as salary deposits, bill payments, retail transactions, and peer-to-peer transfers. After the typologies were assigned, transactions were simulated based on the predefined patterns of each typology. This involved determining the sender and receiver accounts, transaction amounts, timestamps, payment methods, and currency types. By assigning varied typologies to different accounts, the dataset captured a wide range of normal banking activities, ensuring realistic and heterogeneous transaction patterns.

To generate suspicious transactions, random 10-digit bank account numbers were first created to represent fraudulent accounts. Each suspicious account was then assigned a specific money laundering typology, ensuring that different accounts exhibited distinct illicit financial behaviours. These typologies included methods such as surfing and layering, each designed to obscure the origin of illicit funds. Once the typologies were assigned, suspicious transactions were generated by structuring financial flows between multiple sender and receiver accounts. The transaction attributes—such as amount, frequency, currency, and counterparties—were adjusted to reflect the characteristics of the assigned typology. After simulating suspicious transactions, all unique bank accounts involved in suspicious transactions were identified. To enhance realism and make detection more challenging, normal transaction typologies were assigned to these suspicious accounts, allowing them to conduct legitimate financial activities alongside illicit ones. This approach imitated real-world money laundering behaviour, where fraudsters mix illegal funds with legitimate transactions to avoid detection. By integrating normal transactions into suspicious accounts, the dataset ensured that fraudulent activity was more difficult to distinguish, creating a more complex and realistic financial dataset.

Once both datasets were generated, they were integrated into a comprehensive synthetic anti-money laundering dataset, combining normal and suspicious transactions to accurately reflect real-world financial activity. This integration ensured that illicit transactions were embedded within legitimate flows, making the dataset more realistic and suitable for training AML detection models.

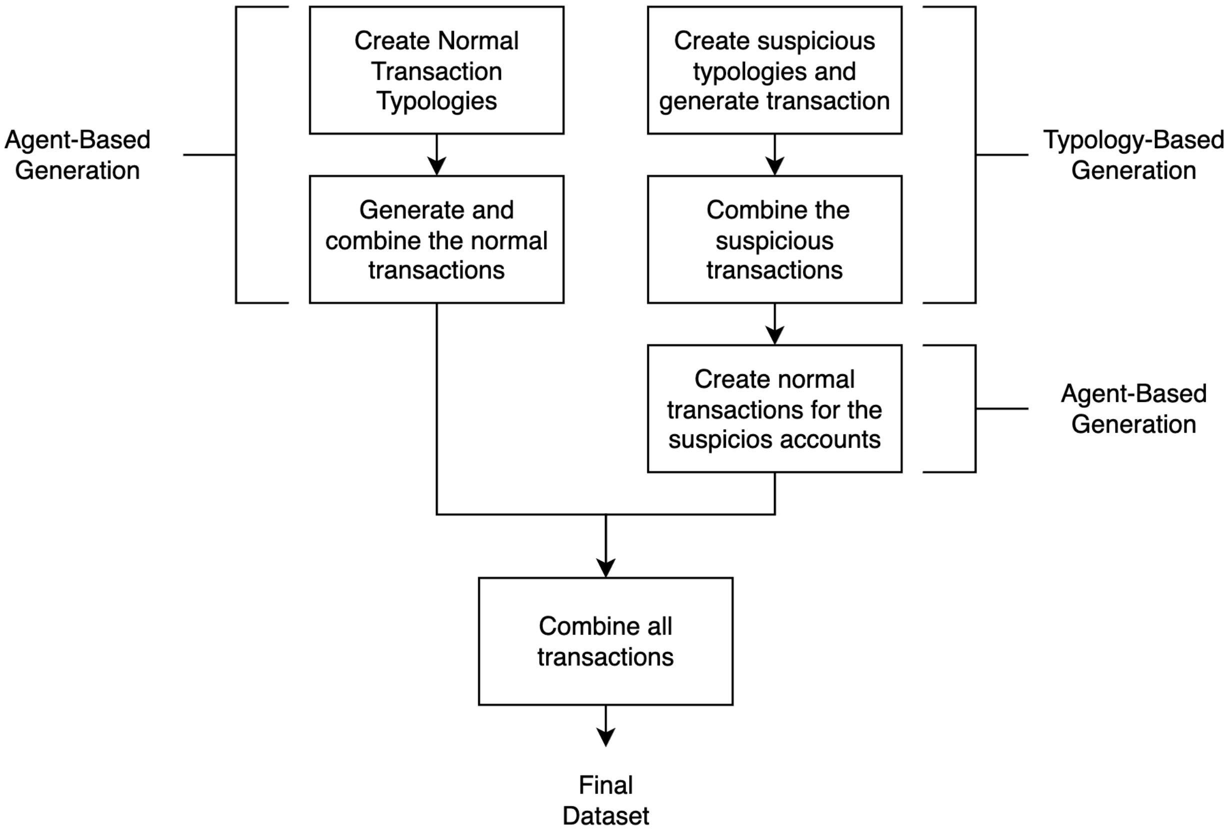

Agent-Based Generation
Create Normal Transaction Typologies
Generate and combine the normal transactions
Create suspicious typologies and generate transaction
Combine the suspicious transactions
Typology-Based Generation
Create normal transactions for the suspicios accounts
Agent-Based Generation
Combine all transactions
Final Dataset

Figure 4.2: Generation process of the 'Normal' and 'Suspicious' accounts and transactions

To represent essential characteristics of banking transactions and include relevant attributes for money laundering transactions, the following includes some of the features comprised in the synthetic dataset; transaction type, transaction amount, sender and receiver bank locations, amongst others. The features were generated through probabilistic modelling to warrant randomness and to reflect the behaviours of real-world banking transactions. Features were chosen precisely through analysing the current AML literature (J. D. J. Rocha-Salazar, M. J. Segovia-Vargas, and M. D. M. Camacho-Miñano, 2021), existing money laundering datasets (Desrousseaux, Bernard, and Mariage, 2021)(Rouhollahi et al., 2021), and attaining specialist input. Notably, the specialist emphasised the critical role of geographic location and payment types, highlighting that transactions involving cross-border transfers and cash payments often exhibit higher risks and warrant closer inspection.

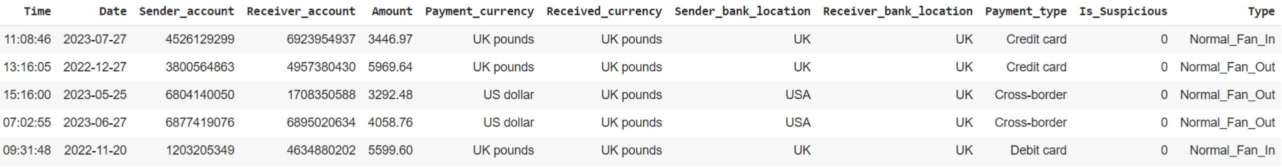
4.4 Dataset Description

A total of 12 features are contained in the dataset which were chosen due to their associations with anti-money laundering transactions. Fig4.3, presents a preview of the generated transactions. The synthetic dataset contains a comprehensive perspective of transactions across multiple banks, focusing on the United Kingdom, as opposed to creating a singular bank viewpoint of the transactions. This broader view of transaction flows enables the measurement of performance enhancements in transaction monitoring, highlighting the potential if banks were to share data. This could be valuable to encourage data sharing amongst financial institutions in the future (Betron, 2012). Creating an individual bank's viewpoint of transactions can be achieved by slightly modifying the generator.

The dataset includes 'Time' and 'Date' features denoting transactional chronology essential to identifying money laundering techniques. 'Sender' and 'Receiver' account details, together with time and date, uncover behavioural patterns and complex banking connections. The 'Amount' feature presents the transaction amounts and can highlight potentially suspicious activities through unusual transaction values. 'Payment Type' represents various transaction methods, each carrying distinct risk levels and regulations. The payment types included in this study's dataset are; credit card, debit card, cash, automated clearing house (ACH) transfers, cross-border, and cheque. Geographic context, inserted through 'Sender Bank Location' and 'Receiver Bank Location', identifies high-risk regions. High-risk countries such as Mexico, Turkey, Morocco, and the United Arab Emirates are included (UK Government, 2020)(Financial Action Task Force (FATF), 2023). 'Payment Currency' and 'Receiver Currency' align with location features, with mismatched instances adding complexity. Finally, the binary 'Is Suspicious' feature differentiates between normal and suspicious transactions, with 'Type' further classifying the typologies, providing deeper insights into prevalent or high-risk transactional typologies.

4.4.1 Typologies

A total of 28 typologies are included in the dataset, split between 11 normal and 17 suspicious. The typologies were chosen based on existing datasets (Suzumura and Kanezashi, 2021), money laundering literature (Simser, 2013), and through semi-structured interviews with AML specialists. The typologies can be expressed using graphical networks to visualise the structure and flow of the transactions. There are 15 different graphical network structures for the 28 typologies. Fig 4.4, presents the 15 different types of structures that the various typologies adopt. Multiple typologies have the same structure to increase complexity, however, the parameters can diverge significantly (i.e. transaction amounts, duration, receiver location, and entities in-



Time	Date	Sender_account	Receiver_account	Amount	Payment_currency	Received_currency	Sender_bank_location	Receiver_bank_location	Payment_type	Is_Suspicious	Type
11:08:46	2023-07-27	4526129299	6923954937	3446.97	UK pounds	UK pounds	UK	UK	Credit card	0	Normal_Fan_In
13:16:05	2022-12-27	3800564863	4957380430	5969.64	UK pounds	UK pounds	UK	UK	Credit card	0	Normal_Fan_Out
15:16:00	2023-05-25	6804140050	1708350588	3292.48	US dollar	UK pounds	USA	UK	Cross-border	0	Normal_Fan_Out
07:02:55	2023-06-27	6877419076	6895020634	4058.76	US dollar	UK pounds	USA	UK	Cross-border	0	Normal_Fan_Out
09:31:48	2022-11-20	1203205349	4634880202	5599.60	UK pounds	UK pounds	UK	UK	Debit card	0	Normal_Fan_In

Figure 4.3: Preview of the generated transactions with features

volved). The typologies with the same structure will have different parameter values that also overlap with each other to increase intricacy, mimic real-world situations, and make detection harder. This section discusses some of the typologies included in the developed dataset, the ones that are selected from the literature, and the typologies produced by the authors.

The AMLSim dataset by IBM proposes 6 normal typologies and 8 AML typologies (Suzumura and Kanezashi, 2021). However, in their simulated dataset, only the following 3 out of the 8 AML typologies are included; Fan-Out, Fan-In, and Cycle. All the proposed normal typologies are included. In the dataset generation method, I included the following suspicious typologies proposed by IBM; Fan- Out, Fan-In, Cycle, Bipartite, Stacked Bipartite, Scatter-Gather, and Gather-Scatter. The normal typologies adopted from IBM in the dataset created in this study are Single Transaction, Fan-Out, Fan-In, Mutual, Forward, and Periodical. The typologies in our dataset adopt the structure of the proposed IBM typologies but increase complexity and randomness by having varying parameter ranges and varying numbers of entities involved each time the typology is generated. The authors further developed the Fan-In and Fan-Out typologies by increasing the transaction layers, thus producing the Layered Fan-In and Layered Fan-Out typologies. These typologies represent the layering process in money laundering (Schönenberg and Schönfeld, 2013). One generation of the Layered Fan-In typologies structure is shown in Fig 4.4, illustrating multiple sender accounts transacting with a fewer number of accounts, who subsequently transact with a single receiver. The Layered Fan-Out typologies mirror this structure but with transactions flowing in the reverse direction. The typologies can accommodate a varying number of accounts within each layer, and the transaction values can increase or decrease by margins of 10% to 20% across each layer depending on the typology.

During the semi-structured interviews, the participants were asked about typologies that are difficult to detect and scenarios that generate high false positives currently in the industry using existing transaction monitoring methods. High-risk transaction types and their characteristics were also discussed, with emphasis put on cash transactions and high-risk geography by the interviewees (M. Riccardi and M. Levi, 2018). Typologies were chosen and created using the results from interviews and the academic literature. The following completes the suspicious typologies included in our dataset; Structuring, Smurfing, Over-Invoicing, Deposit-Send, Cash Withdrawal, Single Large Transaction, Behaviour Change 1, and Behavioural Change 2. Each of these typologies is explained in detail below, providing insights into their characteristics and relevance to money laundering detection. The newly incorporated normal typologies consist of Cash Withdrawal, Cash Deposit, Small Fan-out, Mutual Plus, and Normal Group. During the interviews, the Smurfing (Starnini et al., 2021) and Structuring (El-Banna, Khafagy, and El Kadi, 2020) typologies were frequently mentioned which are considered the most prevalent techniques in the literature concerning the placement stage of money laundering (Irwin, Choo, and L. Liu, 2012)(Unger and Busuioc, 2007). The suspicious Cash Withdrawal typology was created in response to the interviews. It was stated that preventing activities like forced sexual servitude poses a challenge in AML, given the difficulty in detecting low-value withdrawals (McDowell and Novis, 2001). The Deposit-Send typology is considered suspicious due to the rapid movement of funds and potentially facilitating terrorism finance (S. Gao et al., 2006). This typology refers to a situation where an account first deposits cash into the bank and then within a short period of time sends it to another account. The transaction amount is generally below the reporting threshold limit, with the second transaction having an increased chance of being sent to a high-risk country. Another challenge institutions currently face, identified during the interviews, is detecting and monitoring changes in the behaviour of customers. Therefore, Behavioural Change 1, Behavioural Change 2, and Normal Group typologies were created. The Normal Group typology entails an account (main account) that regularly transacts with another group of accounts. The group of accounts is split into two, core accounts and regular accounts. The main account engages in transactions with the core accounts more frequently than the regular accounts. The Behavioural Change 1 and 2 typologies adopt the same structure as the Normal Group typology. However, in Behavioural Change 1, the main account deviates from its usual patterns and transacts with new accounts. In contrast, under the Behavioural Change 2 typology, the main account transacts with new accounts in high-risk locations.

4.5 Results

The proposed generator and typologies were used to create a synthetic transactional dataset for transaction monitoring. The parameters of the dataset were drawn from various sources to enhance credibility and practicality. The knowledge and experience of an AML specialist's input was the main source, which played a fundamental role in deciding the parameters for the normal and suspicious typologies. The criteria for setting the parameters were also determined from assessing the AML literature (Naheem, 2016) and existing transaction datasets (Suzumura and Kanezashi, 2021)(Altman et al., 2023). In this section, I compare the developed dataset to publicly available AML datasets and perform statistical analysis.

4.5.1 Comparison

This section presents the analysis and comparison of datasets specifically designed for money laundering research. These datasets comprise of AMLSim, IT-AML, MLDP, and the newly proposed Synthetic Anti-Money Laundering Dataset (SAML-D). Tables 4.1 and 4.2 provide a comparison of the datasets.

The AMLSim, IT-AML, and SAML-D datasets all include a practical and large amount of transactions. They incorporate a realistic ratio of money laundering to normal transactions, producing an imbalanced dataset that represents real-world circumstances accurately. Additionally, these datasets provide flexibility by allowing for an adjustable volume of transactions during the generation phase. Conversely, the MLDP dataset offers a comparatively limited scope with only 2340 transactions. Moreover, due to an unrealistic proportion of money laundering to normal transactions, the MLDP dataset deviates from what is typically considered real-world contexts. The IT-AML and SAML-D datasets both include various types of transactions, hence offering a more complete perspective than other datasets in comparison. For instance, AMLSim and MLDP include one and two transaction types respectively, presenting a more simplified view of transaction dynamics. The SAML-D dataset includes 28 typologies offering a richer and more nuanced scope compared to the 14 in IT-AML, 9 in AMLSim, and 3 in MLDP. The explanations for each typology are available in the SAML-D, IT-AML, and AMLSim datasets, facilitating a deeper understanding of the data. Whereas, in the MLDP dataset no explanation of the types of money laundering activities is given, though they are labelled within the data. Likewise, the SALM-D dataset labels each typology, enhancing the

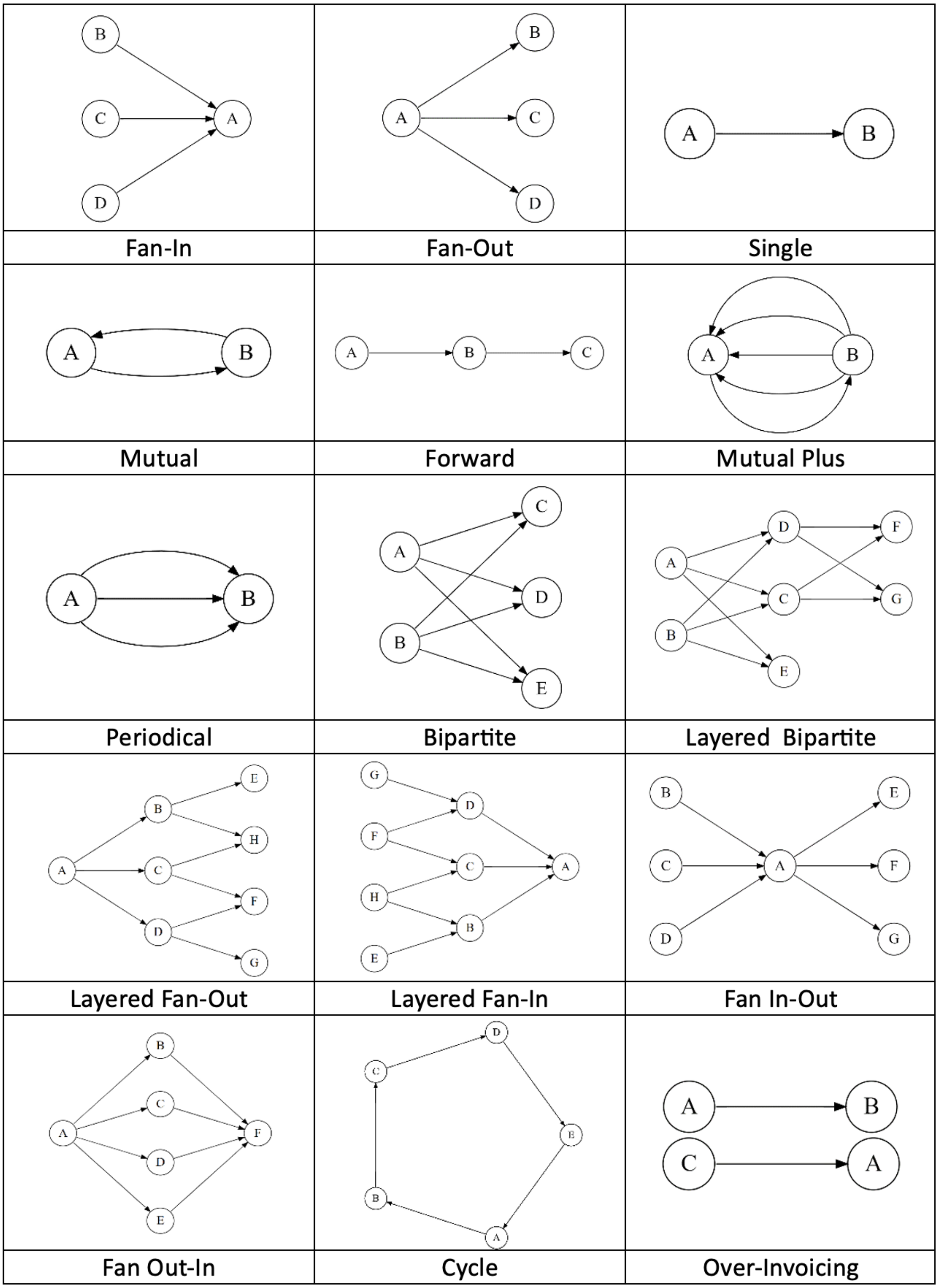


Figure 4.4: Graphical structures of different typologies

interpretability of the dataset meaning a higher accuracy and performance can be reached. The typology label feature is not present in either AMLSim or IT-AML. A unique characteristic

Table 4.1: Comparison of different transaction datasets

Dataset	No. of Features	No. of Transactions	No. of Accounts	No. of SAR transactions	Transactions
AMLSim	8	12 476 012	100 000	17052 (0.137%)	1 - Transfer
PaySim	11	6 362 620	6 353 307	8213 (0.129%)	5 - Cash-in, Cash-out, Debit, Payment, Transfer
IT-AML	11	5 078 345	515 000	5177 (0.102%)	7 - Credit Card, ACH, Wire, Cheque, Cash, Bitcoin, Reinvestment
Credit Card Fraud Detection	31 (anonymized)	284 807	N/A	484 (0.172%)	N/A
MLDP	7	2340	2340	1399 (60%)	2 - Transfer, Cash-in
SAML-D	12	9 411 384	749 507	11658 (0.124%)	6 - Credit Card, Debit Card, ACH, Cheque, Cash, Cross border

of the IT-AML and SAML-D datasets is the incorporation of transactions involving multiple currencies, which adds a layer of complexity and a higher degree of realism to these datasets. The SAML-D dataset further enhances its value and realism through the addition of geographic locations, even featuring high-risk countries. This feature aligns the dataset more closely with features often employed in the industry.

Overall, the proposed dataset, SAML-D, provides a detailed and robust resource for transaction monitoring in the field of AML. SAML-D contributes to the AML domain by introducing innovative features and typologies facilitating effective and complex analysis of transaction monitoring approaches.

4.5.2 Experiment

In this section I conduct machine learning experiments on the newly developed dataset, SAML-D, and the AMLSim dataset, to detect suspicious transactions. The AMLSim dataset serves as a benchmark due to its frequent utilisation in the AML literature. By applying identical

Table 4.2: Comparative analysis of key characteristics across transaction datasets

	AMLSim	PaySim	IT-AML	Credit Card Fraud Detection	MLDP	SAML-D
Suspicious Typology Explanation	Yes (3)	No	Yes (8)	No	No	Yes (17)
Normal Typology Explanation	Yes (6)	No	Yes (6)	No	No	Yes (11)
Model Multiple Currencies	No	No	Yes	N/A	No	Yes
Model Geographic Locations	No	No	No	N/A	No	Yes
Labelled Typologies	Yes	No	No	No	No	Yes

machine learning algorithms with consistent hyperparameters across both datasets, these experiments aim to clarify which dataset is more challenging for accurately identifying suspicious transactions, thereby highlighting the relative complexity and usefulness of SAML-D within AML research.

The selected machine learning algorithms are chosen to capture various characteristics and complexities within each dataset, including Support Vector Machines (SVM), a distance-based model adept at handling complex class boundaries; Naïve Bayes (NB), a probabilistic model; Decision Trees, specifically the Classification and Regression Tree (CART) model ideal for understanding feature importance and their interplay; and Random Forest, an ensemble model. The models were trained using the sklearn library, with all hyperparameters set to their default values. These methods were chosen as they are some of the most utilised techniques in the literature (Berkan Oztas et al., 2022)(Suresh and Padmajavalli, 2006). Ultimately, these experiments seek to establish the suitability, utility, and practical relevance of the newly developed SAML-D dataset relative to the widely utilised AMLSim dataset.

Data analysis was the first step to get a better understanding of the data, identify the data types for each feature, and check for missing values. Next, the data was pre-processed, transforming it into a suitable format for the chosen machine learning algorithms. Pre-processing involved converting categorical features, such as payment type, into numerical form by the use of one-hot encoding and label encoding. This allowed the algorithms to interpret the information more effectively (Suresh and Padmajavalli, 2006). The date feature was split into year, month, and day, while the is fraud feature in AMLSim was converted from a boolean datatype to an integer. After redundant columns were dropped, both datasets were standardised re-scaling the numerical variables to a comparable scale. This was especially important for the transaction amount feature, given its excessive range. Also, some algorithms like SVM are sensitive to the scale of the input features and therefore perform better on standardised data.

A 70-15-15 stratified train-validation-test split was adopted for the experiments, implemented as a single time-based split to mimic real-world scenarios. This approach ensured that the temporal order of the data was preserved, reducing data leakage risks. The 70% training data enabled to effectively fit the model to the data, while the 15% validation data was utilised for preliminary performance evaluation. Finally, the 15% test data was used to evaluate the models. This approach helped prevent model overfitting and enabled generalised results. Due to the constraint of a single GPU, the experiments were conducted on a representative subset of the datasets.

The evaluation metrics used to assess the performance of the models include the True Positive Rate (TPR), True Negative Rate (TNR), False Positive Rate (FPR), False Negative Rate (FNR), and ROC-AUC score. These metrics are widely used in the AML literature and provide valuable insights into the model's performance (Alsuwailem and Saudagar, 2020). The TPR measures the proportion of actual suspicious transactions that are correctly identified. Assessing the TPR is crucial, as missing suspicious transactions can result in significant consequences for banks in the form of fines and potential legal action by authorities. The TNR indicates the percentage of normal transactions that are correctly labelled as normal. The evaluation metrics also assessed the model's error. The FPR quantifies the transactions that are incorrectly labelled as suspicious by the model, which can lead to high operational costs and wasted resources for banks. On the contrary, the FNR presents the rate of suspicious transactions that are mislabelled as normal and go undetected by the institution. A high FNR can result in reputational damage and regulatory fines. The ROC-AUC score represents the model's overall performance and efficiency in classifying between normal and money laundering transactions. In summary, the chosen evaluation methods provide both an overview of the model's performance and a nuanced analysis of its specific elements.

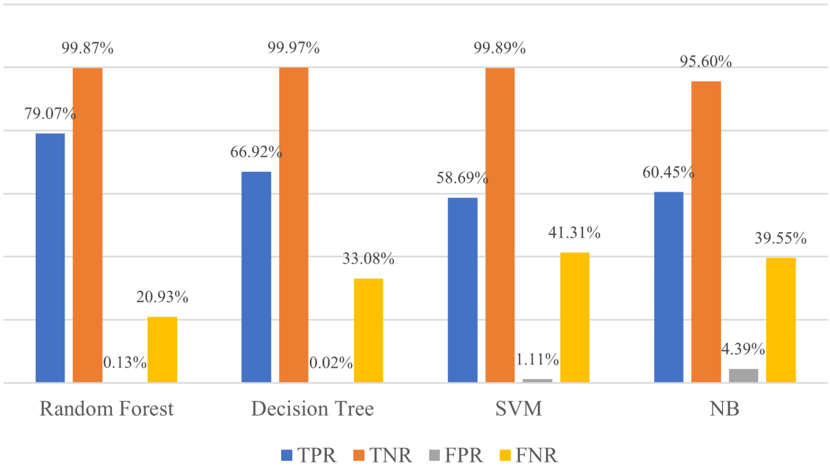


Figure 4.5: Experiments results for the SAML-D dataset

All the model's performances had a lower TPR and higher FNR on the SAML-D dataset compared to the AMLSim dataset, as shown in Fig 4.5 and Fig 4.6. The results indicate that these models encountered challenges in detecting money laundering transactions in the SAML-D dataset, implying a higher level of complexity or deception in the money laundering structures and patterns. Despite the lower FPR observed in the AMLSim dataset, the models successfully detected most of the money laundering transactions. This capability is crucial and

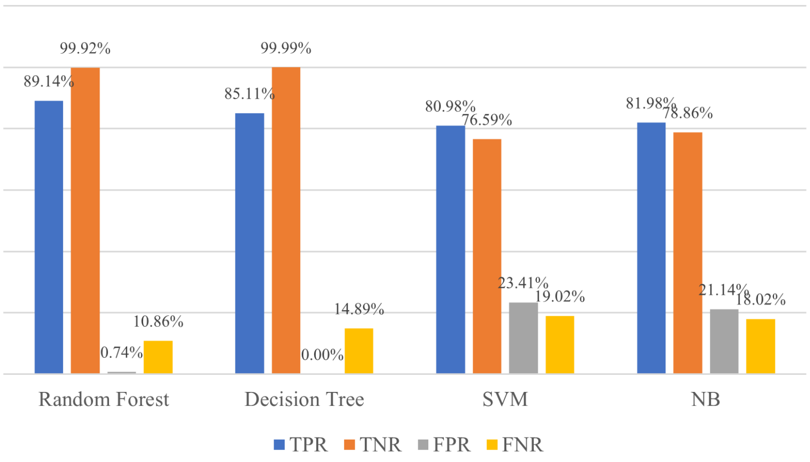


Figure 4.6: Experiments results for the AMLSim dataset

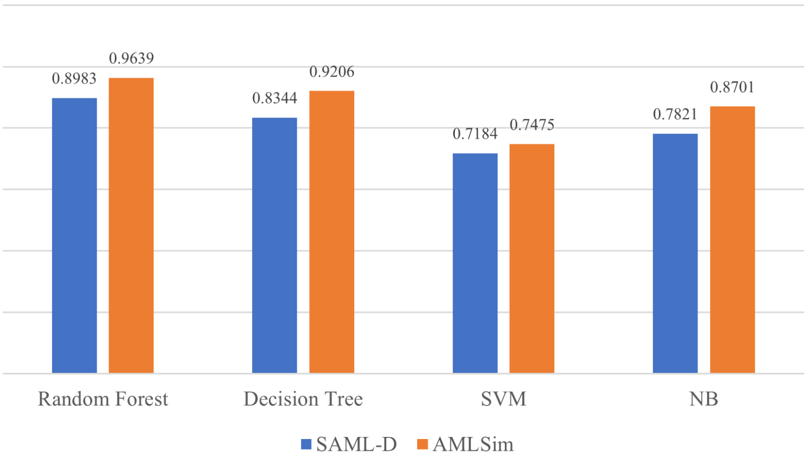


Figure 4.7: Comparison of ROC-AUC score for the SAML-D and AMLSim datasets

of greater importance for financial institutions than having a model with a higher FPR, as failure to detect suspicious behaviours could potentially lead to fines and reputational damage. The ROC-AUC score presents a similar trend with the SAML-D dataset attaining a lower score for every model, shown in Fig 4.7. This emphasises the difficulty these machine learning models had in detecting suspicious transactions in the SAML-D dataset. The findings demonstrate that while these machine learning models perform well on the AMLSim dataset their performance weakens on the SAML dataset. This highlights the need for sophisticated and robust models. Potential approaches could involve feature engineering to produce more predictive attributes and exploring more advanced or tailored machine learning models.

4.6 Summary

In conclusion, this research addresses the challenge of accessing AML transaction monitoring data, which is typically unavailable due to legal and privacy constraints, or limited in terms of true labels and diversity. The novel synthetic AML transaction generator provides a valuable resource to advance transaction monitoring in the field of AML. Using the generator a dataset called SAML-D was created. The primary purpose of the generator and dataset is to provide researchers with an additional resource to evaluate their models and facilitate a comparative analysis of their results.

The SAML-D dataset contains 12 features and 28 typologies, including both 'normal' and 'suspicious' entities. The typologies were chosen and created based on existing datasets, the literature, and 8 semi-structured interviews with AML specialists. This study's approach brings significant enhancements compared to other synthetic datasets such as AMLSim, IT-AML, and MLDP. Key enhancements include the introduction of geographic locations and the attention given to high-risk payment types, giving the dataset a high degree of realism that reflects the complexities seen in real-world industry situations. Additionally, the SAML-D dataset includes a vaster selection of typologies offering a richer and more nuanced scope compared to the other datasets.

In testing the SAML-D dataset against the more established AMLSim dataset through machine learning experiments, I found that models had more difficulty identifying suspicious transactions within the SAML-D dataset. This implies a higher level of complexity and difficulty in the money laundering patterns and structures, making SAML-D a valuable resource for future studies in this domain.

Despite its contribution, it is vital to recognise that the SAML-D dataset has limitations. As it is a synthetic dataset, it will not fully capture the intricacy and unpredictability of real-world transactions. Some parameters, while based on informed estimations and expert consultations, may not embody all real-world scenarios. Moreover, the typologies included will not encapsulate all possible money laundering strategies, particularly due to the criminal's ever-changing techniques. However, the SAML-D dataset offers a resource for researchers and practitioners to conduct experiments and compare their findings, potentially assisting the development of more advanced and capable transaction monitoring methods.

As for future work, I aim to conduct experiments with more complex machine learning algorithms on the SAML-D dataset, leading to further improvements in detecting suspicious activities.

Chapter 4 References

Alsuwailem, Abdullah A.S. and Akhtar Jamal Khan J. Saudagar (2020). "Anti-money laundering systems: a systematic literature review". In: *Journal of Money Laundering Control* 23.4, pp. 833–848. DOI: 10.1108/JMLC-02-2020-0018.

Altman, E. et al. (2023). *Realistic Synthetic Financial Transactions for Anti-Money Laundering Models*. arXiv preprint. Available from: https://arxiv.org/abs/2306.16424 [Accessed 7 Mar. 2024].

El-Banna, M. M., M. H. Khafagy, and H. M. El Kadi (2020). "Smurf Detector: a Detection technique of criminal entities involved in Money Laundering". In: *2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE)*. Aswan, Egypt, pp. 64–71.

Betron, M. (2012). "The state of anti-fraud and AML measures in the banking industry". In: *Computer Fraud & Security* 2012.5, pp. 5–7.

Chen, S. H. and R. Venkatachalam (2017). “Agent-based modelling as a foundation for big data”. In: *Journal of Economic Methodology* 24.4, pp. 362–383.

Cheng, X. et al. (2021). “Combating emerging financial risks in the big data era: A perspective review”. In: *Fundamental Research* 1.5, pp. 595–606.

Desrousseaux, R., G. Bernard, and J. J. Mariage (2021). “Profiling Money Laundering with Neural Networks: a Case Study on Environmental Crime Detection”. In: *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 364–369. ISBN: 2375-0197. DOI: 10.1109/ICTAI52525.2021.00059.

Europol (2017). *From Suspicion to Action: Converting Financial Intelligence into Greater Operational Impact*. Europol. Available from: https://www.europol.europa.eu/cms/sites/default/files/documents/ql-01-17-932-en-c_pf_final.pdf [Accessed 9 Mar. 2024].

Financial Action Task Force (FATF) (2023). *High-risk and Other Monitored Jurisdictions - June 2023*. Available from: https://www.fatf-gafi.org/en/publications/High-risk-and-other-monitored-jurisdictions/Increased-monitoring-june-2023.html [Accessed 12 Mar. 2024].

Gao, S. et al. (2006). “Intelligent Anti-Money Laundering System”. In: *2006 IEEE Int. Conf. Serv. Oper. Logist., Informatics*. Shanghai, China, pp. 851–856.

He, Ping (2010). “A typological study on money laundering”. In: *Journal of Money Laundering Control* 13.1, pp. 15–32.

Irwin, S. M., A. Raymond Choo, and L. Liu (2012). “Modelling of money laundering and terrorism financing typologies”. In: *Journal of Money Laundering Control* 15.3, pp. 316–335.

Johnson, M. C. and S. R. Kessler (2019). “The art and science of semi-structured interviewing: A comprehensive guide for researchers”. In: *Qualitative Research Journal* 21, pp. 131–147. DOI: 10.1177/1468794119825569.

Jullum, Martin et al. (2020). “Detecting money laundering transactions with machine learning”. In: *Journal of Money Laundering Control* 23.1, pp. 173–186. ISSN: 1368-5201. DOI: 10.1108/JMLC-07-2019-0055.

Le Borgne, Y.A. et al. (2022). *Reproducible Machine Learning for Credit Card Fraud Detection – Practical Handbook*. Université Libre de Bruxelles. Available from: https://github.com/Fraud-Detection-Handbook/fraud-detection-handbook [Accessed 16 Mar. 2024].

Lopez-Rojas, E. A. and S. Axelsson (2012). “Money laundering detection using synthetic data”. In: *Annual Workshop of the Swedish Artificial Intelligence Society (SAIS)*. Linköping University: Linköping University Electronic Press.

Mahootiha, M. (2020). *Money Laundering Data*. Kaggle. Available from: https://www.kaggle.com/datasets/maryam1212/money-laundering-data [Accessed 05 Feb. 2024].

McDowell, John and Gary Novis (2001). “The consequences of money laundering and financial crime”. In: *Economic Perspectives* 6.2, pp. 6–10.

Naheem, Mohammed Ahmad (2016). “Money laundering: A primer for banking staff”. In: *International Journal of Disclosure and Governance* 13.2, pp. 135–156. ISSN: 1746-6539.

Oztas, Berkan et al. (2022). “Enhancing Transaction Monitoring Controls to Detect Money Laundering Using Machine Learning”. In: *2022 IEEE International Conference on e-Business Engineering (ICEBE)*, pp. 26–28. DOI: 10.1109/ICEBE55470.2022.00014.

Plaksiy, K., A. Nikiforov, and N. Miloslavskaya (2018). "Applying big data technologies to detect cases of money laundering and counter financing of terrorism". In: *2018 6th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*. IEEE, pp. 70–77.

Riccardi, M. and M. Levi (2018). "Cash, Crime and Anti-Money Laundering". In: *The Palgrave Handbook of Criminal and Terrorism Financing Law*. Ed. by C. King, C. Walker, and J. Gurulé. Cham: Palgrave Macmillan.

Rocha-Salazar, J. D. J., M. J. Segovia-Vargas, and M. D. M. Camacho-Miñano (2021). "Money laundering and terrorism financing detection using neural networks and an abnormality indicator". In: *Expert Systems with Applications* 169, pp. 1–15. DOI: 10.1016/j.eswa.2020.114470.

Rouhollahi, Z. et al. (2021). "Towards Proactive Financial Crime and Fraud Detection through Artificial Intelligence and RegTech Technologies". In: *The 23rd International Conference on Information Integration and Web Intelligence*, pp. 538–546. DOI: 10.1145/3487664.3487740.

Schönenberg, Regine and Annette von Schönfeld (2013). *Transnational organized crime: Analyses of a global challenge to democracy*. transcript Verlag. DOI: 10.14361/transcript.9783839424957.

Shokry, Amr Ehab Muhammed, Mohammed Abo Rizka, and Nevine Makram Labib (2020). "Counter terrorism finance by detecting money laundering hidden networks using unsupervised machine learning algorithm". In: *International Conference on e-Learning*, pp. 89–97. DOI: 10.33965/ict_csc_wbc_2020_2020081012.

Simser, J. (2013). "Money laundering: emerging threats and trends". In: *Journal of Money Laundering Control* 16.1, pp. 41–54.

Starnini, M. et al. (2021). "Smurf-Based Anti-money Laundering in Time-Evolving Transaction Networks". In: *Machine Learning and Knowledge Discovery in Databases. Applied Data Science Track*. Vol. 12978. Lecture Notes in Computer Science. Springer. DOI: 10.1007/978-3-030-86514-6_11.

Suresh, R. M. and R. Padmajavalli (2006). "An overview of data preprocessing in data and web usage mining". In: *2006 1st Int. Conf. Digit. Inform. Manage.* IEEE, pp. 193–198.

Suzumura, T. and H. Kanezashi (2021). *Anti-Money Laundering Datasets: InPlusLab Anti-Money Laundering Datasets*. IBM GitHub Repository. Available from: http://github.com/IBM/AMLSim [Accessed 9 Aug. 2024].

Tundis, A., S. Nemalikanti, and M. Mühlhäuser (2021). "Fighting organized crime by automatically detecting money laundering-related financial transactions". In: *The 16th International Conference on Availability, Reliability and Security*. Vol. 38, pp. 1–10. DOI: 10.1145/3465481.3469196.

UK Government (2020). *National Risk Assessment of Money Laundering and Terrorist Financing 2020*. UK Government. Available from: https://www.gov.uk/government/publications/national-risk-assessment-of-money-laundering-and-terrorist-financing-2020 [Accessed 22 May. 2024].

Unger, B. and E.M. Busuioc (2007). *The Scale and Impacts of Money Laundering*. Edward Elgar, Cheltenham, UK; Northampton, MA. Available from: http://www.loc.gov/catdir/toc/ecip074/2006035573.html [Accessed 29 Mar. 2024].

Valbuena, D., P. H. Verburg, and A. K. Bregt (2008). “A method to define a typology for agent-based analysis in regional land-use research”. In: *Agriculture, Ecosystems & Environment* 128.1, pp. 27–36.

Chapter 5

Transformer Based Transaction Monitoring Model: Tab-AML

Version of record available at: B. Oztas, D. Cetinkaya, F. Adedoyin, M. Budka, H. Dogan and G. Aksu, "Tab-AML: A Transformer Based Transaction Monitoring Model for Anti-Money Laundering" in 2025 IEEE Conference on Artificial Intelligence (CAI), Santa Clara, California, USA. .
DOI: 10.1109/ICEBE59045.2023.00028

5.1 Introduction

This chapter builds on inputs from Chapters 2, 3, and 4 to experiment with deep learning models in the transaction monitoring domain, addressing objective (iv). Figure 5.1 illustrates the research motivation from Chapter 2, the insights gained from Chapter 3, and the application of the dataset generator from Chapter 4, all of which contributed to the development of this chapter. As part of this integrated thesis, this chapter is intended for submission to a peer-reviewed journal.

This study proposes a deep learning approach to conduct transaction monitoring. The objective is to enhance upon rule-based methods, by reducing the number of false positive alerts while maintaining a high true positive rate, utilising machine learning techniques. This study's focus is to explore the application of transformers (Vaswani et al., 2017) to enhance transaction monitoring, an area largely unexplored in this context. The significant contributions of the research include the application of existing transformer-based models to the transaction monitoring domain and the enhancement of a transformer-based model specifically designed for this task. I conduct in-depth experiments to compare the performance of the transformer-based model (Tab-AML), the known TabTransformer model (X. Huang et al., 2020), the TabNet model (Arik and Pfister, 2021), and the various baseline models. Hyperparameter tuning is employed to optimise the model's results. The results provide an understanding of how Tab-AML performs compared to the TabTransformer, TabNet, and baseline machine learning techniques that are widely used in the literature for transaction monitoring. This research highlights the performance of Tab-AML and demonstrates how transformer-based models can

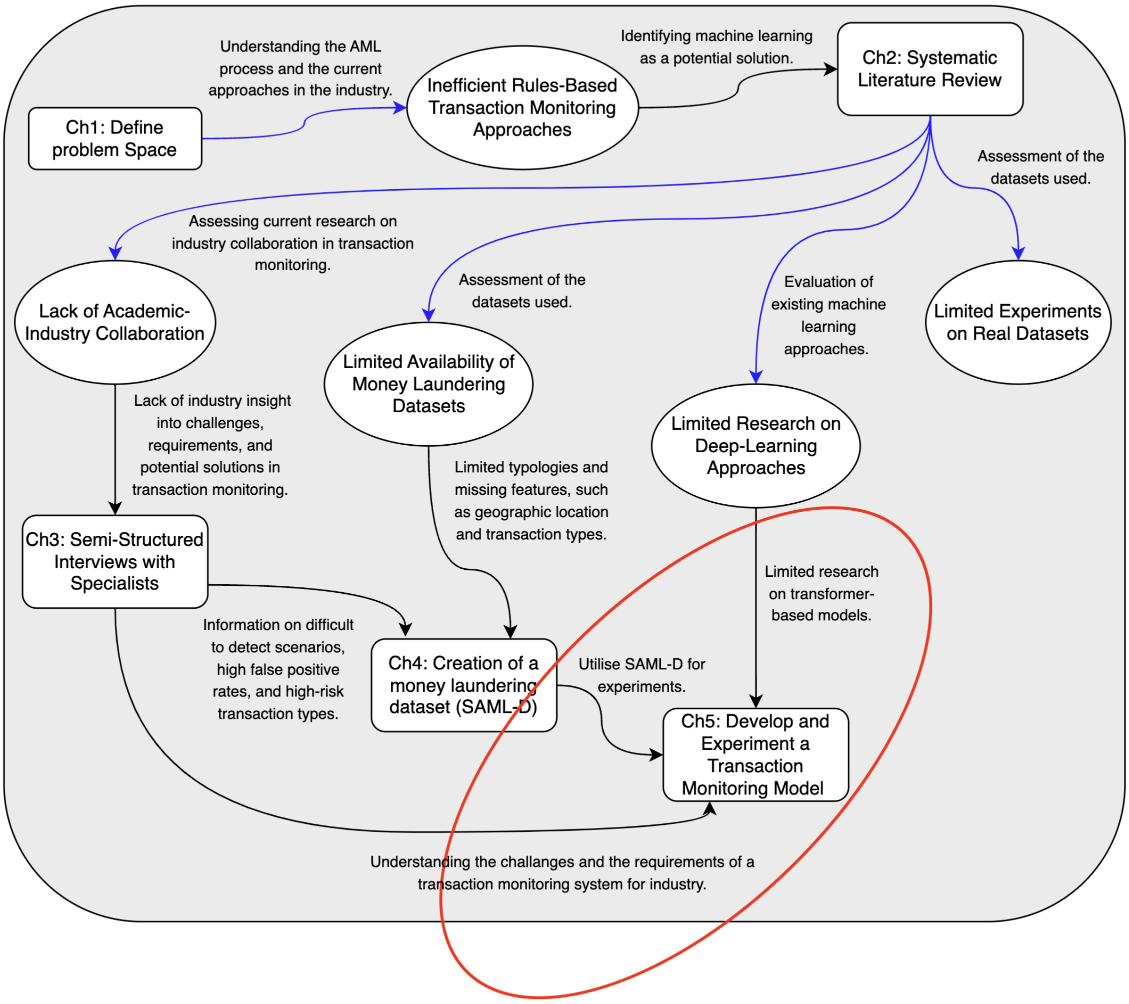

Ch1: Define problem Space
Understanding the AML process and the current approaches in the industry.
Inefficient Rules-Based Transaction Monitoring Approaches
Identifying machine learning as a potential solution.
Ch2: Systematic Literature Review
Assessment of the datasets used.
Assessing current research on industry collaboration in transaction monitoring.
Lack of Academic-Industry Collaboration
Assessment of the datasets used.
Limited Availability of Money Laundering Datasets
Evaluation of existing machine learning approaches.
Limited Experiments on Real Datasets
Limited Research on Deep-Learning Approaches
Lack of industry insight into challenges, requirements, and potential solutions in transaction monitoring.
Limited typologies and missing features, such as geographic location and transaction types.
Ch3: Semi-Structured Interviews with Specialists
Limited research on transformer-based models.
Information on difficult to detect scenarios, high false positive rates, and high-risk transaction types.
Ch4: Creation of a money laundering dataset (SAML-D)
Utilise SAML-D for experiments.
Ch5: Develop and Experiment a Transaction Monitoring Model
Understanding the challanges and the requirements of a transaction monitoring system for industry.

Figure 5.1: Overview of the PhD (extracted from Figure 1.3)

potentially enhance the transaction monitoring process for financial institutions.

The structure of this chapter is as follows. Section 5.2 presents a review of the related literature. The dataset used for the experiments is introduced and discussed in Section 5.3. Section 5.4 outlines the architecture of Tab-AML, the transformer-based transaction monitoring model developed in this study. Section 5.5 focuses on the inner workings of the attention and residual attention mechanisms. Section 5.6 discusses the experimental set up including the evaluation metrics, baseline machine learning models, and the hyperparameter tuning process. In Section 5.7 the results are presented and discussed. Finally, Section 5.8 concludes the chapter by summarising the findings.

5.2 Related Work

Machine learning has been utilised by many researchers to detect money laundering in various domains. This review primarily focuses on studies that specifically apply machine learning to develop transaction monitoring systems, aimed at detecting suspicious transactions in financial institutions. Additionally, I expand this review to include studies that employ machine learning to identify money laundering activities in other fields, using different types of data.

A range of techniques has been used in the literature for identifying money laundering transactions, such as logistic regression, decision trees, random forests, neural networks (NN), and support vector machines (SVM) (Y. Zhang and Trubey, 2019). To prevent overfitting risks, the authors utilised the 'nnet' package in R and assessed these models using ROC-AUC curves, sampling strategies, visual representations, and regression analysis. The NN model, specifically the single hidden layer feedforward network, outperformed the rest of the models attaining more desirable results. Similarly, study (Lv, Ji, and J. L. Zhang, 2008) developed a transaction monitoring method using a neural network variant known as radial basis function (RBF). This method employed APC-III clustering to calculate the parameters for the hidden layer in the RBF model, improving the learning speed. To update the weights between the different layers the recursive least square (RLS) algorithm was used over the gradient decent algorithm, as it enhanced convergence speed and performance. Their results showed that the developed RBF model outperformed an SVM and an outlier detection method in terms of true and false-positive rates.

Another study in the literature also experimented with and compared a variety of machine learning methods for transaction monitoring (Tundis, Nemalikanti, and Mühlhäuser, 2021). The following approaches were applied and evaluated; decision trees, SVM, linear regression, gaussian naive bayes, and random forests. To conduct the experiments, a synthetic dataset was used to train and evaluate the performance of the models. To compare the models, accuracy, recall, precision, and F1- score were the chosen evaluation metrics. Their findings revealed that the random forest model outperformed the others in classifying money laundering transactions. The authors conducted further analysis, identifying the most influential attributes for the random forest and decision tree models. The authors in (Ketenci et al., 2021) developed a transaction monitoring method using random forest, focusing on time-frequency attributes of customer transactions. This method was trained on pre-labelled data and set a threshold for flagging suspicious transactions. The study emphasised the usefulness of time-frequency characteristics in simplifying attribute selection and improving performance.

The Isolation Forest and One-Class SVM techniques were investigated by (Amr Ehab Muhammed Shokry, Mohammed Abo Rizka, and Nevine Makram Labib, 2020) to detect money laundering transactions. The authors used a synthetic dataset focusing on the transaction amounts and timings to highlight the suspicious activities. Isolation forest proved to attain a higher accuracy with a faster computation time and less memory usage. Another study, (Guevara, Garcia-Bedoya, and O. Granados, 2020), presented the significance of the isolation forest method in inspecting withdrawal information of transactions. A real dataset with nine features was used, potentially offering a more accurate real-world representation.

A study proposed using the XGBoost method for detecting suspicious transactions (Jullum et al., 2020), outperforming a bank's existing system in metrics like ROC-AUC, Brier score, and their developed PPP metric. Their approach included transactional alerts from rule-based methods that didn't lead to AML reports, to improve the results and representativeness. Additionally, (Stojanović et al., 2021) compared Random Forest, AdaBoost, and XGBoost across three datasets, with XGBoost performing best in two. They focused on sensitivity and specificity but did not account for the false-positive rate, an important metric for financial institutions.

An Autoencoder and Variational Autoencoder were presented by (Zhiyuan et al., 2021) to enhance transaction monitoring. Their models solely focused on normal transactions during training and aimed to detect suspicious transactions by assessing recreation errors when testing. A novel approach was generating a mixed dataset, including synthetic transactions, created using the Wasserstein Generative Adversarial Network (WGAN), which reduced the false positive rate from 19% to 8%. Despite this improvement, challenges regarding overfitting and low precision were identified.

A Self-Organising Map (SOM) with added clustering was proposed by (Alshantti and Rasheed, 2021) for transaction monitoring. This method categorises neurons into risk-level clusters based on inter-neural distance. The experiment was conducted on a small dataset (by DNB Bank) including real and labelled transactions. While it significantly reduced the false positive rate to 6.2%, the true positive rate was only 65.5%, indicating a risk of overlooking actual money laundering activities. The study noted the method's decreased efficiency with imbalanced data, a common matter in transaction monitoring.

The following discussion focuses on machine learning methods developed for detecting money laundering activities, although they are not specifically tailored for transaction monitoring in financial institutions. An autoencoder to identify fraud and money laundering in Brazilian export data was employed in (Paula et al., 2016). The model utilises 80 various attributes provided from the data to learn patterns and excels in identifying anomalous situations. They used the mean squared error reconstruction error as a metric to determine suspicious activities. Out of the 819,990 Brazilian firms in the dataset, the study identified 20 high-risk firms, of which some were confirmed to be fraudulent by third-party experts.

Another approach utilised in the field of AML is graph neural networks. This study, (Alarab, Prakoonwit, and Nacer, 2020) investigated Graph Convolutional Networks (GCNs) to detect illicit Bitcoin transactions on the blockchain. The authors develop a novel approach combining GCNs with linear layers and a multi-layer perceptron. The model was evaluated on an elliptic dataset and attained desirable results. The developed approach outperformed methods that relies only on traditional GCNs or Skip-GCNs, achieving better precision (0.899), recall (0.678), F1 (0.773), and accuracy (0.974) scores.

A transformer-based approach to detect money laundering in capital markets called HAMLET was created by the authors in (Tatulli et al., 2023). It utilises a hierarchical transformer architecture to identify complex patterns and behaviours. A synthetic dataset including 29,704,090 capital market transactions performed by 400 financial institutions is used to evaluate the model. The model achieves a 99% precision score in binary classification and 95% in multi-class classification. The dataset included five different money laundering schemes simulating real-world capital market scenarios.

While these studies demonstrate the potential of various machine learning techniques in detecting money laundering, significant research gaps remain. Notably, there is a lack of studies employing deep learning methods specifically for transaction monitoring within the AML domain (Kute et al., 2021). Furthermore, although transformer-based models have shown promise in other contexts, such as capital markets, their application to transaction monitoring in financial institutions remains under explored. Addressing these gaps is essential to improve the efficiency and performance of AML systems, particularly in reducing false positives and ensuring comprehensive detection of illicit activities across different financial domains.

5.3 Data

In this work, I employed the dataset named SAML-D (B. Oztas et al., 2023a), a synthetic dataset developed for transaction monitoring and calibrated with expert insights from an industrial collaborator. The creation of SAML-D involved, semi-structured interviews with AML specialists, analysis of existing datasets, and a comprehensive review of the relevant literature. The SAML-D dataset contains 9,504,852 transactions, including 0.104% suspicious transactions equivalent to precisely 9,873 entries. A total of 676,912 unique sender and receiver accounts engage in transactions, encompassing 18 distinct geographic locations, utilising 13 varied currencies, and facilitated through 7 different types of payment methods. The dataset's 28 typologies, including 11 normal and 17 suspicious types, are modelled to reflect real-world scenarios. The suspicious typologies involve various money laundering types such as layering techniques (i.e., structuring), suspicious behavioural changes over time, rapid movement of funds, and transactions to high-risk geographical locations, amongst many others. An overlap in suspicious and normal typologies increases the complexity of the dataset making it harder to identify the money laundering transactions. A summary of the 12 features in the original SAML-D dataset is provided in Table 5.1, before pre-processing.

Feature	**Description**
Temporal details	Date and time stamps of transactions.
Account details	The account number of the sender and receiver of the transaction.
Transaction Amount	Indicating the monetary value of each transaction.
Bank Locations	Geographical data points for the sender and receiver accounts.
Currency Information	Currency of the transaction for the sender and receiver accounts.
Payment Method	Types of payments like wire transfer, credit, debit, etc.
Suspicion Indicator	A binary value signaling normal or suspicious transactions.
Typology Classification	Identifying the specific AML typology of each transaction.

Table 5.1: SAML-D Features

5.3.1 Categorical Encoding

Categorical attributes such as geographical location, payment type, and currency are converted to numerical format. I considered various encoding techniques, including Binary Encoding, One Hot Encoding, and Label Encoding. The Label Encoding approach was selected due to its simplicity by offering a more compact representation of the categorical data while maintaining performance. This approach transforms each category within each feature by assigning an increasing integer value. For example, encoding the geographic locations: “UK” as 0, “USA” as 1, “Turkey” as 2, and so on. Although Label Encoding is typically suited for ordinal data, in the transformer model's, due to the utilisation of embedding layers the allocated integers hold no practical value. The embedding layers can effectively interpret the categorical data regardless of the assigned integers, hence, not impacting the model performance.

5.3.2 Log Transformation

Addressing data skewness is a significant step in pre-processing the data, specifically the *amount* feature that indicates the monetary value of each transaction. The *amount* feature values present substantial variability, with a large skew to the right, shown in Figure 5.2. Most transaction amounts are clustered at lower values, with others at considerably larger values. Skewed data can negatively impact a model's performance, as many algorithms perform best with normally distributed inputs on which to base their predictions (Tanha et al., 2020). Additionally, standardising the skewed data before handling the issue would result in misrepresenting the original data's distribution. Log transformation addresses the problem and normalises the distribution of the data by compressing the range of higher amounts and expanding the lower range. The influence of outliers is also minimised on the model, which is crucial given the wide range of transaction amounts. The result of applying log transformation is a well-represented dataset, ensuring more accurate and reliable model predictions.

The log transformation is defined as follows:

$$X = \log(y + 1) \tag{5.1}$$

where X denotes the post-log transformation values and y the pre-transformed values.

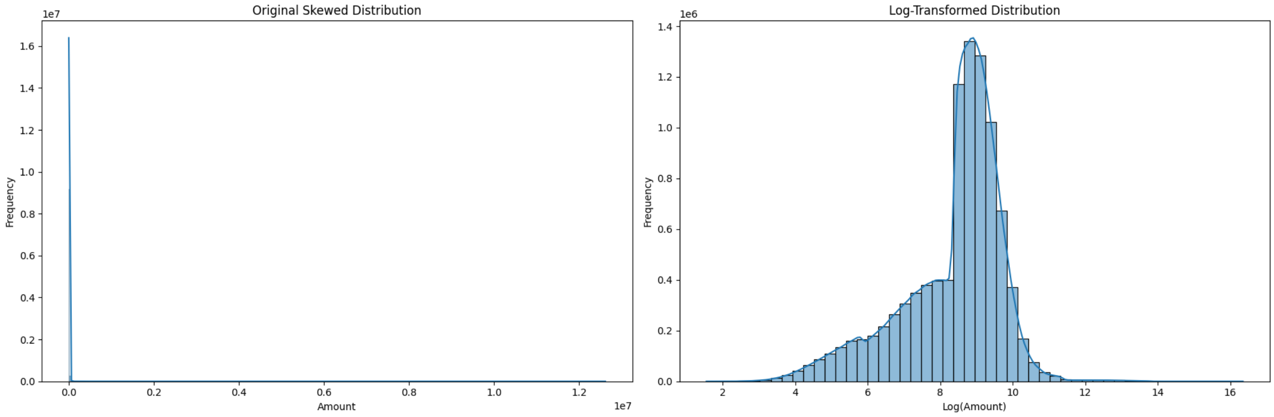


Figure 5.2: The original and log-transformed distributions of transaction amounts.

5.3.3 Standardisation

Standardisation was applied to the non-categorical features within the dataset to improve the convergence and efficiency of the models (Gal and Rubinfeld, 2019). This standardisation process, executed after the log transformation step, aligns the data around a mean of zero and a standard deviation of one. I calculated the mean (μ) and standard deviation (σ) for standardisation using only the training dataset, hence avoiding any data leakage from the validation and test data sets, maintaining model integrity. Also, standardisation was executed on each feature independently, as in the following equation,

$$Z = \frac{X - \mu}{\sigma} \tag{5.2}$$

where Z denotes the standardised values. While I focused exclusively on standardisation, exploring alternative methods such as Min-Max Scaling might offer different perspectives on model performance. However, given the nature of money laundering detection models, where preserving the original distances between data points is crucial, standardisation is the preferred choice in this study's implementation.

5.3.4 Feature Engineering

The dataset's depth was developed further by generating new features from the 'Date' and 'Time' attributes. The 'Date' feature was used to derive the 'Day', 'Month', and 'Year' components, while the 'Time' feature facilitated the creation of the 'Hour' feature. Additionally, I generated the 'Is Weekend' and 'Day of Week' features. Table 5.2 provides the features with descriptions.

Feature	**Description**
Day	Day of the transaction relative to the month.
Month	Month of the transaction.
Year	Year of the transaction.
Hour	The hour the transaction occurred.
Is Weekend	If the transaction occurred over the weekend.
Day of Week	The day of the week the transaction occurred.

Table 5.2: Transaction Time Features

These additions enable more detailed temporal analysis and support the model in recognising patterns regarding specific time frames. This approach is important for identifying subtle patterns in transaction behaviour that may indicate suspicious activity, hence increasing the effectiveness of the models. Finally, I drop the 'Date' and 'Time' columns along with the 'Typology Classification' feature as it will not be used in the experiments.

5.4 Model Architecture

In this section, I present an in-depth overview of the Tab-AML architecture, illustrated in Figure 5.3. The primary objective of Tab-AML is to accurately predict and identify suspicious transactions that may require further investigation and potential reporting to higher authorities. The motivation for developing this model stems from the limitations of current rule-based methods used by institutions, which often fail to effectively classify suspicious transactions and frequently produce a significant number of false positives. Additionally, this study explores the application of transformers for transaction monitoring within the realm of Anti-Money Laundering (AML). To the best of my knowledge, the use of transformers specifically for transaction monitoring is largely unexplored. This research builds upon the TabTransformer model (X. Huang et al., 2020), a deep learning approach tailored for tabular data.

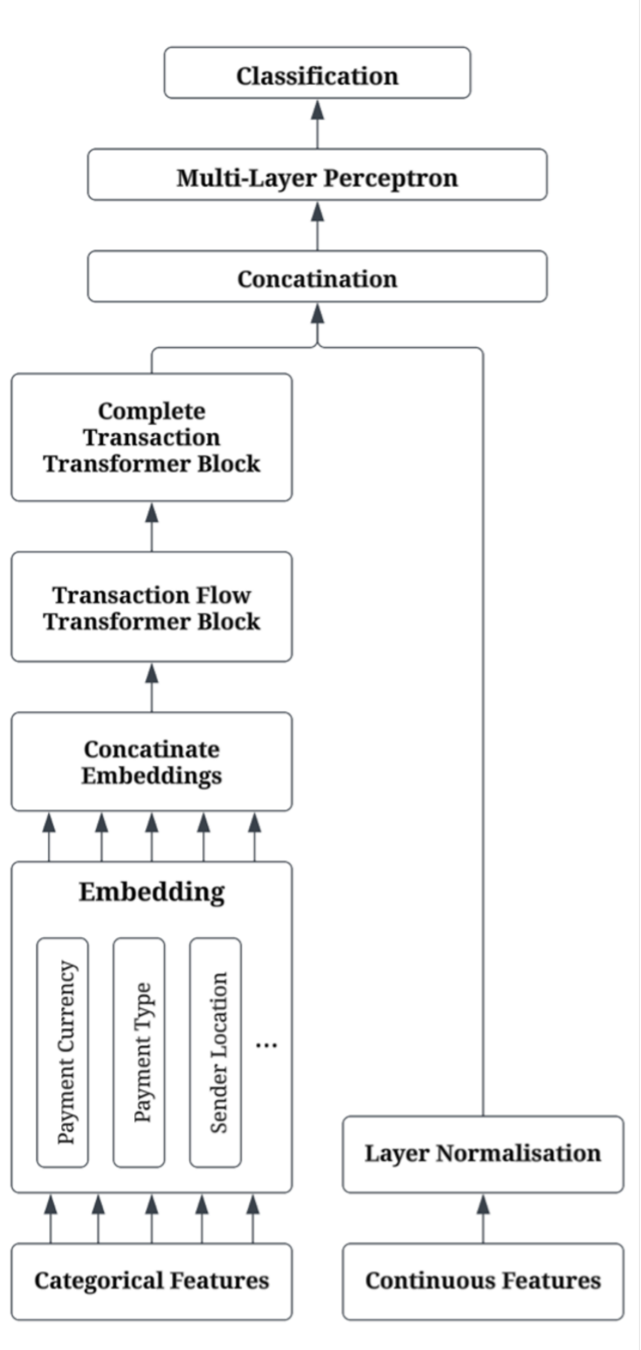


Figure 5.3: Overview of the Tab-AML Models' Flow.

5.4.1 Embedding

This stage involves the embedding of categorical features and preparing the continuous features. This was crucial for the utilised dataset since it included both feature types, a common characteristic of transactional datasets. Categorical features are individually embedded using the Column embedding approach. The embedding dimensions are set to match the model's dimensions (d), adhering to the method used in X. Huang et al. This uniformity simplifies the architecture by removing the need for adjustments between layers (Dahouda and Joe, 2021). Additionally, the embedding layers are initialised with a truncated normal distribution (mean of zero and standard deviation of one), chosen to stabilise the learning process by limiting excessively large initial weights that could negatively impact training (Narkhede, Bartakke, and

Sutaone, 2022). To identify the most appropriate embedding size, I explored dimensions of 16, 32, 64, and 128. Detailed insights and results from these experiments are elaborated in Section 5.6. In total 12 features are embedded and their unique counts are presented in Table 5.3.

In addition, the Tab-AML model utilises a shared embedding mechanism to further enhance the embedding process. This shared embedding component consists of a universal embedding vector that is concatenated to the individual embeddings of each feature. This design enables a portion of the embedding vector to be shared across different features, which can help the model generalise better across various features. The embedding matrix for individual features is specifically tailored to accommodate the shared component by reducing its dimensionality accordingly. This adjustment ensures that when the shared embedding vector is concatenated to each feature's individual embedding, the original dimensionality is preserved, maintaining a uniform feature representation across the model. I explored various ratios for the shared embedding size relative to the individual embeddings, specifically fractions of 1/4, 1/8, and 1/12 of the total embedding dimension. After testing, I chose a fraction of 1/8, as it yielded the best performance.

Following the embedding process, all categorical feature embeddings are concatenated along the feature dimension, aligning them side-by-side within the same tensor. This concatenated vector allows the model to assess interactions among features, uncovering underlying patterns within the dataset.

While the Tab-AML model utilises uniform embedding dimensions aligned with the model's dimensionality, future explorations could investigate varying embedding dimensions specifically tailored to the unique characteristics of each feature. This approach could potentially uncover more complex nuances in features with high cardinality.

Table 5.3: Overview of Embedded Features and Their Unique Counts

Feature	**Unique Counts**
Sender Account	292,715
Receiver Account	652,266
Payment Currency	13
Received Currency	13
Sender Bank Location	18
Receiver Bank Location	18
Payment Type	8
Date - Year	2
Date - Month	11
Date - Day	31
Hour	24

5.4.2 Transformers

The concatenated embeddings are then fed into the transformer encoder blocks (Vaswani et al., 2017). The Tab-AML model, created for money laundering detection, incorporates two sequential encoders with residual attention mechanisms to process the feature set strategically. The first encoder processes the complete feature set, applying a masked attention mechanism that exclusively focuses on the 'Sender Account' and 'Receiver Account'. This selective attention aims to effectively isolate and enhance the interaction analysis between these pivotal features, crucial in contexts where money laundering techniques often involve complex transaction networks designed to hide the illicit origins of funds, such as through structuring (O. M. Granados and Vargas, 2022). The output from the first encoder, comprising both enhanced and standard feature representations, serves as the input to the second encoder. In this stage, no masking is applied, allowing the second encoder to deploy its attention mechanism across all the features. This unmasked processing integrates the interactions learned from the 'Sender Account' and 'Receiver Account' with the broader dataset context, outputting contextual embeddings. These embeddings are created from a combination of focused initial learning and broad feature integration, aiming to capture detailed relationships crucial for the model's accuracy and its ability to generalise.

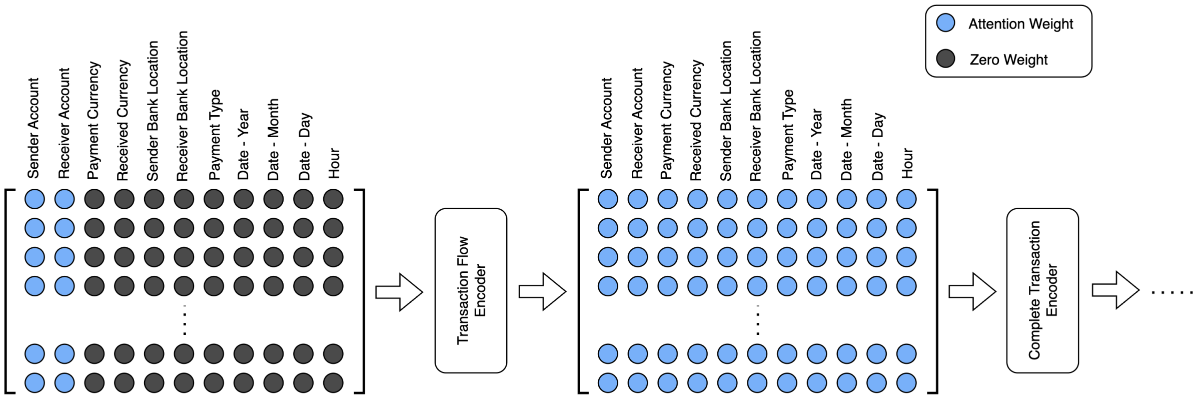


Figure 5.4: Dual-masked transformer encoder structure highlighting selective feature processing for money laundering.

Residual attention is incorporated due to its capacity to enhance the models' focus on relevant data points while retaining context from previous layers (R. He et al., 2021). This feature can be influential for money laundering detection, as understanding the relationship between current and past transactions is essential. This approach utilises sparse attention-like features, allowing it to focus on critical information to identify suspicious activities. Additionally, the residual attention contributes to greater stability and enhanced regularisation within the model. This reduces the risk of overfitting to specific patterns and behaviours, improving the ability to generalise across various laundering techniques. Further details on the residual attention can be found in Section 5.5, and is presented in Figure 5.5. A residual dropout is also applied (Srivastava et al., 2014).

5.4.3 Classification

In the final stages of Tab-AML's architecture, the outputted contextual embeddings are concatenated with the continuous features. This concatenation results in a vector of dimensions $((d \times num\_cat) + num\_con)$, where d represents the embedding dimension, num_cat is the number of categorical features, and num_con is the number of continuous features. This vector is then inputted into a Multi-Layer Perceptron (MLP) to classify the transactions (Hervé and Jean-François, 2018).

The MLP layer sizes are determined by the multiplication factors specified in the hyperparameters, which are set to (4, 2), as in X. Huang et al. This configuration first expands the input dimension (l) by a factor of 4, leading to a layer size of 4 $\times$ l in the first hidden layer. Subsequently, the dimension is reduced by a factor of 2 in the second hidden layer, resulting in a size of 2 $\times$ l. The final output layer compresses the processed features into a single scalar value, using a sigmoid activation function to output a probability indicating the likelihood of a transaction being associated with money laundering (LeCun et al., 2012). The hidden layers employ the Gaussian Error Linear Unit (GELU) activation function, offering a smooth, non-linear transformation, critical for identifying complexities in the data (Hendrycks and Gimpel, 2016). To prevent overfitting and enhance the model's generalisation, a dropout rate is applied within the MLP.

5.5 Attention and Residual Attention Mechanism

In this section, I introduce the transformer encoder (Luong, Pham, and Manning, 2015), (Vaswani et al., 2017), specifically the scaled dot-product attention and multi-head attention, as they are key components of the Tab-AML models architecture.

A fundamental component of the transformer is the scaled dot-product attention, which uses the following set of vectors: queries (Q), keys (K), and values (V), to calculate attention (Vaswani et al., 2017). A general attention mechanism can be described by the following formula:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{5.3}$$

The calculations of Q, K, V, and d_k are presented below.

- The input data is used to derive the queries, keys, and values. Let X be an input matrix. Then $Q = XW^Q, K = XW^K$, and $V = XW^V$, where W^Q, W^K, and W^V are weight matrices specific to queries, keys, and values, respectively.
- d_k is the dimensionality of the key and query vectors.

Once the necessary matrices are derived, the dot product of the query with all the keys is computed, resulting in a matrix of scores. The scores are then scaled down by the square root of d_k, to stabilise the gradient during training. The Softmax function is applied to the updated scores to obtain and normalise the attention weights, ensuring they sum up to 1. The attention weights are then multiplied with the value vectors to produce the output of the attention layer.

The multi-head attention mechanism expands on the single attention by running multiple scaled dot-product attention instances in parallel (Vaswani et al., 2017). This enhances the model to process and integrate multiple representation subspaces, leading to better learning capabilities and performances. The multi-head attention formula is:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_n)W^O \tag{5.4}$$

Where, $\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$. The matrices W_i^Q, W_i^K, and W_i^V are responsible for linearly projecting queries, keys, and values into the attention space of the i-th head. The W^O matrix then linearly transforms the concatenated outputs from all the attention heads.

Each transformer layer also incorporates a fully connected Feed-Forward Network (FFN) that has a single hidden layer, defined as:

$$\text{FFN}(x) = \sigma(xW_1 + b_1)W_2 + b_2. \tag{5.5}$$

The activation function is denoted by σ. In the Tab-AML model, σ corresponds to the GELU activation function. This FFN is applied independently and identically to each position within the sequence. Additionally, layer normalisation modules are integrated above the multi-head attention and FFN to stabilise training (Ba, Kiros, and Hinton, 2016).

5.5.1 Residual Attention

Residual attention is an addition to the standard attention mechanism with enhanced abilities to capture long-range dependencies more effectively. It achieves this by incorporating the attention score from the previous layer, before Softmax, into the current layer's attention calculation (R. He et al., 2021). This modification is shown in Figure 5.5 and can be described by the following formula:

$$\text{Attention}_{\text{res}}(Q, K, V, \mathit{prev}) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}} + \mathit{prev}\right) V \tag{5.6}$$

Where Q, K, and V are the queries, keys, and values, respectively, as in standard attention and *prev* represents the attention scores from the previous layer. The rest of the steps align with the standard attention mechanism.

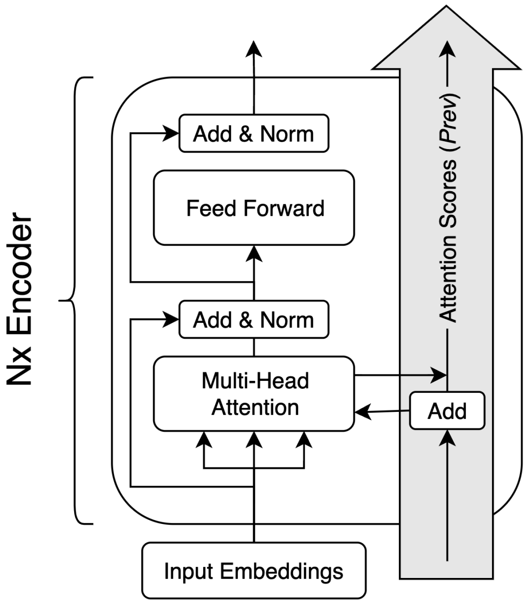


Figure 5.5: Transformer Encoder with residual attention used in Tab-AML

5.6 Experimental Settings

Our experiments aim to compare the Tab-AML model against frequently employed techniques in money laundering detection, as well as against the TabTransformer and TabNet models. Additionally, I focus on optimising the performance of the models on the dataset through hyperparameter tuning. The optimal parameters identified for each model are then chosen and applied to a holdout dataset, allowing me to evaluate the performance for detecting suspicious transactions. Tables 5.6 and 5.7 provide the optimal performance of each model on the validation dataset (with final parameters) and the test dataset.

An 80-10-10 time based train-validation-test split was adopted for the experiments, differing from the split used in Chapter 4 due to the availability of a larger dataset and more computational power. Prior to this split, the transactions were chronologically ordered based on their date and time. The training set, comprising 80% of the data, consists of 7,603,881 transactions and is utilised to fit the model to the data. The validation and test datasets equally share the remaining 20% of the data, each including 950,485 transactions. The validation dataset is used for hyperparameter tuning and the test dataset for evaluating the model's capabilities. I closely monitored for signs of overfitting using the validation and test datasets results, which is when the model learns the training dataset to an extent where it generalises negatively, impacting the model's performance on new datasets. I considered retraining the model on a combined training and validation dataset after hyperparameter tuning but opted against it. This was to prevent introducing any bias into the model's performance.

5.6.1 Evaluation Metrics

To evaluate and compare the various models, I selected the ROC-AUC score (Fawcett, 2006) as the evaluation metric. As the dataset is highly imbalanced, a typical characteristic of money laundering data, the traditional accuracy metric is unsuitable. The aim is to evaluate the model's capability to differentiate and identify between money laundering and normal transactions. Therefore, utilising the accuracy metric would be insufficient for this purpose. The ROC curve visualises the performance of models by plotting the rate of true positives against the rate of false positives across a range of decision thresholds. The area under the ROC curve, called the ROC-AUC score, provides a single gauge of the model's discriminative ability. This metric ranges from 0.5 to 1, where a score of 0.5 represents a model equivalent to randomly guessing, and a score of 1 implies a perfectly discriminative model. Once the model with the best performance is identified I will analyse the results of that model further by plotting the ROC curve and producing a confusion matrix at a given threshold. The confusion matrix describes a model's performance by presenting the counts of truly and falsely predicted positives and negatives. The adopted evaluation metrics are widely used in the AML literature and offer valuable insights into the model's abilities (Alsuwailem and Saudagar, 2020). I will also examine the convergence curves of both training and validation ROC-AUC scores, along with loss scores, to understand the models' learning dynamics over time. These convergence curves offer critical insights into how quickly and effectively each model adapts to the training data, as well as how it generalises to the validation data.

5.6.2 Hyperparameter Tuning: Deep-Learning Models

This section outlines the hyperparameter tuning strategy, detailing the exploration of specific value ranges and the final values chosen for the models. Default values were used for any hyperparameters not explicitly discussed. Due to computational constraints, I opted for a manual and targeted hyperparameter strategy instead of more resource-intensive methods like GridSearch or Optuna (Akiba et al., 2019). This approach involved systematically testing one hyperparameter at a time, keeping others constant to isolate the effects of individual hyperparameters. This study's efforts aimed to maximise the ROC-AUC score on the validation dataset, thus enhancing both model performance and generalisability. Despite the limitations of manual tuning, such as time intensity and the potential for incomplete exploration, this method yielded valuable insights into hyperparameter sensitivity and model behaviour. Future research with more resources may explore alternative optimisation methods to improve efficiency and coverage.

To ensure a comprehensive representation of the model's performance with the final chosen hyperparameter values, I repeated experiments three times using different seed values. The ROC-AUC score was averaged across these experiments and used as the final evaluation metric, including the standard deviation. An early stopping mechanism, utilising the validation dataset's ROC-AUC score, was implemented to optimise training. For the Tab-AML and Tab-Transformer models, training was halted if there was no significant enhancement in the score above a threshold of $1e^{-4}$. For the TabNet model, the training process stopped if there was no improvement in the ROC-AUC score for five consecutive epochs.

Tab-AML and TabTransformer

To maximise the performance of the Tab-AML and TabTransformer models, an exploration of hyperparameters was conducted. I based the hyperparameter settings on the recommendations and findings detailed in X. Huang et al., 2020 and R. He et al., 2021, ensuring alignment with established research practices. Both models employed the Binary Cross-Entropy with Logits Loss (BCEwithLogitLoss) as their loss function due to its inherent numerical stability. This function, which integrates Sigmoid activation with Binary Cross-Entropy Loss (BCE Loss), offers enhanced reliability over the standard BCE Loss.

To optimise the models' performance, I tested a range of learning rates (lr), from $1e^{-5}$ to $1e^{-2}$, to determine the most effective rates for balancing learning speed with convergence accuracy. The learning rate was set at $1e^{-4}$ for the Tab-AML model and $1e^{-3}$ for the Tab-Transformer model. I evaluated the efficacy of various optimisers (O), including Adam, AdamW (Llugsi et al., 2021), and Stochastic Gradient Descent (SGD) (Bottou, 2012), all implemented with their default parameters. Ultimately, the AdamW optimiser was selected for both models as it slightly outperformed the Adam optimiser.

In terms of model dimensionality (d), values of 16, 32, 64, and 128 were tested. The selected dimension significantly influenced the models' learning capacity and data embedding size, impacting their overall efficiency. Notably, higher dimension values increased model complexity, which led to overfitting, while excessively low values yielded suboptimal outcomes. Consequently, a dimension of 32 was identified as optimal for the Tab-AML model, and 64 for the TabTransformer model, providing a balanced approach to maximising both efficiency and performance.

Further experiments were conducted on the encoder's multi-head attention mechanism within each model, evaluating configurations of 4 heads with 2 layers and 8 heads with 4 layers. The configuration of 4 heads with 2 layers was found to offer the best generalisability and ROC-AUC score, indicating that while additional heads and layers might capture a broader range of dependencies, they did not significantly enhance performance and could potentially lead to overfitting.

Batch size (B) was another influential variable, with sizes of 32, 64, 128, and 256 assessed. Larger batch sizes facilitated quicker convergence but at the cost of reduced generalisation. Conversely, smaller batch sizes improved model outcomes but required longer training durations. However, a significantly small batch size lead to overfitting. An optimal batch size of 64 was selected for the Tab-AML model and 128 for the TabTransformer model, offering an effective balance between training efficiency and convergence stability.

To prevent overfitting and enhance generalisation in the Tab-AML and TabTransformer models, I implemented dropout in various components (Srivastava et al., 2014). For the Tab-AML model, a residual dropout rate of 10% was selected, while the TabTransformer model utilised a 10% attention dropout. Additionally, Tab-AML employed a 10% MLP dropout rate and TabTransformer had a 10% feed-forward dropout. These rates were determined to be optimal after testing 0%, 10%, and 20%. Selecting a 10% rate effectively reduced overfitting while preserving essential learning capacity, ensuring robust performance.

Table 5.4: Tested Hyperparameters for the Tab-AML and TabTransformer Models

Parameter	**Tested Values**
Learning Rate (lr)	$1e^{-5}$, $1e^{-4}$, $1e^{-3}$, $1e^{-2}$
Optimisers (O)	Adam, AdamW, SGD
Model Dimensionality (d)	16, 32, 64, 128
Batch Size (B)	32, 64, 128, 256
Multi-Head Attention Configuration	4 heads with 2 layers, 8 heads with 4 layers
Residual/Attention Dropout	0%, 10%, 20%
MLP/Feed Forward Dropout	0%, 10%, 20%

TabNet

I optimised the TabNet model's hyperparameters to maximise the ROC-AUC score. I followed the hyperparameter guidelines in the TabNet paper (Arik and Pfister, 2021) and utilised a PyTorch implementation of the model. I used cross-entropy loss as the loss function and the Adam optimiser with its default settings. The depth of the decision steps (N_d) and the width of the attention layers (N_a) were both tested with values of 16, 32, and 64, with 32 emerging as the optimal setting for both parameters. This gave an efficient trade-off between performance and complexity, preventing overfitting risks. I experimented with 1, 3, and 5 as possible values for the number of decision steps (N_{steps}), identifying that a setting of 1 step was most effective. A low N_{steps} allowed for a well generalised model. The feature scaling parameter (γ) was varied among 1.0, 1.5, and 2.0, with the highest value of 2.0 being selected. Sparse regularisation (λ_{sparse}) was assessed using the following values, $1e^{-6}$, $1e^{-4}$, and $1e^{-2}$. The smallest value of

$1e^{-6}$ was chosen allowing the model to use more features. The values explored for the batch size (B) were 2048, 4096, and 16384, with 4096 found as optimal. For the virtual batch size (VB), used for ghost batch normalisation (Hoffer, Hubara, and Soudry, 2017), the values tested were 512, 1024, and 2048. A value of 1024 was identified as the best option. The momentum (M) settings of 0.1, 0.25, and 0.4 were explored, with the highest value of 0.4 adopted. The learning rate (lr) was examined at $1e^{-2}$, $2e^{-2}$, $3e^{-2}$, with $2e^{-2}$ selected as it provided the best convergence rate without causing the training to diverge. Future studies might explore a gradually decaying learning rate to potentially enhance convergence and performance.

Table 5.5: Tested Hyperparameters for the TabNet Model

Parameter	**Tested Values**
Learning Rate (lr)	$1e^{-3}$, $1e^{-2}$, $1e^{-1}$
Depth of Decision Steps (N_d)	16, 32, 64
Width of Attention Layers (N_a)	16, 32, 64
Number of Decision Steps (N_{steps})	1, 3, 5
Feature Scaling (γ)	1.0, 1.5, 2.0
Sparse Regularisation (λ_{sparse})	$1e^{-6}$, $1e^{-4}$, $1e^{-2}$
Batch Size (B)	2048, 4096, 16384
Virtual Batch Size (VB)	512, 1024, 2048
Momentum (M)	0.1, 0.25, 0.4

5.6.3 Hyperparameter Tuning: Baseline Models

To get an in-depth understanding of the model's performance, I selected a range of baseline machine learning techniques alongside the TabTransformer model for comparison. The techniques were chosen based on their frequent use for money laundering detection in the literature (Niveen M. Labib, Mai A. Rizka, and Ahmed E. M. Shokry, 2020; Alsuwailem and Saudagar, 2020; Berkan Oztas et al., 2022; Lokanan, 2022), offering a diverse view for analysing the model's effectiveness. However, due to the confidentiality and unavailability of the datasets used in previous studies, direct comparison was not possible. Hence, I experimented with the baseline machine learning techniques on the dataset to ensure a meaningful comparison. Each baseline model was optimised using GridSearchCV (P. B. Liashchynskyi and P. Liashchynskyi, 2019) to conduct hyperparameter tuning, with CV set to 3 and scoring based on the ROC-AUC score. This approach assisted in evaluating the model but also contributed to a more in-depth understanding of the selected machine learning techniques to detect money laundering patterns. Implementation of these techniques mainly involved utilising the Scikit-learn library Pedregosa et al., 2018. The exception was Extreme Gradient Boosting (XGBoost) T. Chen and Guestrin, 2016, which was implemented using the XGBoost library. Any hyperparameters not explicitly mentioned were kept at their default library values. Below I present the chosen baseline approaches:

1. **Logistic Regression** (Bisong, 2019), as applied by Sudjianto et al., 2010, is tested using the lbfgs solver with L2 Regularisation Strength values of $C \in \{0.01, 0.1, 1, 10, 100\}$.

2. **K-Nearest Neighbours (KNN)** (Mucherino, Papajorgji, and Pardalos, 2009), previously utilised by Hampo, Nwokorie, and Odii, 2023, is evaluated with a range of neighbours: $N \in \{2, 3, 5, 7, 9\}$.

3. **Decision Tree, (CART)** (Quinlan, 1986), referenced by R. Liu et al., 2011, are explored for Maximum Tree Depth values of $D \in \{2, 4, 8, 16\}$ and Minimum Samples Split, $S \in \{4, 8, 12\}$.

4. **Random Forest** (Breiman, 2001), as cited by Tundis, Nemalikanti, and Mühlhäuser, 2021, is tested for Maximum Tree Depth value of $D \in \{2, 4, 8, 12\}$ and Maximum Features to Consider per Split, $F \in \{2, 4, 8, 12\}$.

5. **XGBoost** (T. Chen and Guestrin, 2016), noted for its application by Jullum et al., 2020, is evaluated for Maximum Tree Depth, $D \in \{2, 4, 8, 16\}$ and Learning Rate options, $L \in \{0.1, 0.2, 0.3\}$.

6. **Gaussian Naive Bayes (GaussianNB)** (Bishop and Nasrabadi, 2006), as highlighted by Lokanan, 2022, is also included to benchmark its effectiveness without hyperparameter tuning.

5.7 Results

Once I identified the optimal parameters with the validation dataset, I retrained the models on the test dataset using the optimised settings. The test dataset, unseen by the models, was used as a benchmark to evaluate and compare the performance of the models. Tables 5.6 and 5.7 highlight the effectiveness of the Tab-AML model in detecting suspicious transactions compared to the other models. Tab-AML achieves an ROC-AUC score of 92.83% on the validation dataset and 93.01% on the test dataset, outperforming the other approaches. The TabTransformer model outperformed TabNet and the baseline models attaining an ROC-AUC score of 87.66% (validation dataset) and 85.94% (test dataset). TabNet attained an ROC-AUC score of 81.78% on the validation dataset and 80.61% on the test dataset. The traditional machine learning approaches, such as Logistic Regression and K-Nearest Neighbours, got significantly lower ROC-AUC scores compared to the transformer-based methods. The XGBoost model performed well outperforming the other traditional methods. This was in line with the existing literature as XGBoost has been shown to outperform other algorithms, such as decision tree and random forest, in identifying suspicious transactions (Domashova and Mikhailina, 2021).

The enhanced performance of Tab-AML can be attributed to the architectural improvements. Tab-AML incorporates a dual-masked Transformer encoder structure with residual attention, enabling comprehensive analysis at both micro and macro levels. Also, a shared embedding mechanism is utilised facilitating better learning of feature interactions, making Tab-AML more effective at handling complex patterns. The first transformer block focuses on transaction flows, analysing interactions between the sender and receiver account features. The second block extends to a holistic examination using the entire feature set, capturing further patterns. An influential attribute setting Tab-AML apart is its residual attention mechanism.

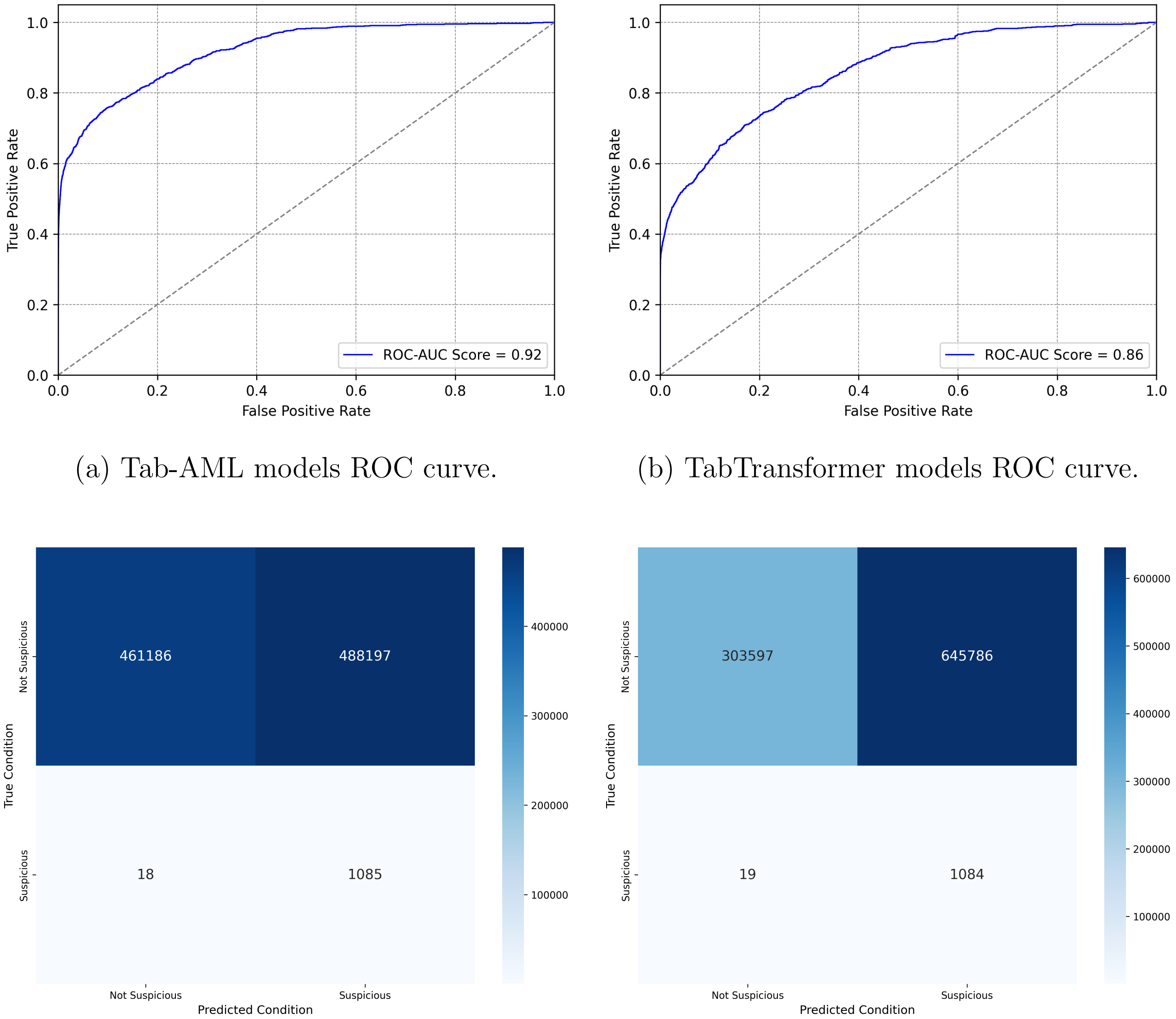
Models	**ROC-AUC Score**	**Hyperparameters**
Tab-AML	**92.83** (0.13)	O = AdamW, $lr = 1e^{-4}$, $d = 32$, $B = 64$, Heads = 4, Layers = 2
TabTransformer	87.66 (0.69)	O = AdamW,lr= $1e^{-3}$, $d = 64$, $B = 128$, Heads = 4, Layers = 2
TabNet	81.78 (0.01)	O = Adam, lr= $2e^{-2}$, $N_d = N_a = 32$, $N_{\text{steps}} = 1$, $\gamma = 2$, $\lambda_{\text{sparse}} = 1e^{-6}$, $M = 0.4$, $B = 4096$, $VB = 512$
Baseline Models		
Logistic Regression	59.74	$C = 100$
K-Nearest Neighbours	59.09	$N = 7$
Decision Tree, (CART)	78.44	$D = 8$, $S = 4$
Random Forest	80.51	$D = 8$, $F = 8$
XGBoost	81.64	$D = 4$, $L = 0.1$
Gaussian Naive Bayes	66.55	-

Table 5.6: ROC-AUC Scores and Best Hyperparameters for the Models on the Validation Dataset. Displays the average ROC-AUC scores, computed over three different seeds, alongside the optimal hyperparameters identified for each model.

Models	**ROC-AUC Score**
Tab-AML	**93.01** (0.61)
TabTransformer	85.94 (0.24)
TabNet	80.61 (0.03)
Baseline Models	
Logistic Regression	56.59
K-Nearest Neighbours	59.61
Decision Tree, (CART)	77.85
Random Forest	79.69
XGBoost	81.12
Gaussian Naive Bayes	63.89

Table 5.7: ROC-AUC Scores of the Models on the Test Dataset. Displays the average ROC-AUC scores for each model, computed over three different seeds.

This mechanism enhances the model's ability to identify long-range dependencies between transactions (R. He et al., 2021). Unlike the traditional attention mechanism, the residual attention mechanism maintains and utilises information over extended batches, improving the detection of complex money laundering patterns involving multiple transactions over time. This leads to better handling of the temporal and relational aspects of transaction data, crucial for identifying suspicious activities. The TabTransformer model outperformed TabNet. Its ability to model complex interactions among categorical features made it well-suited for the study's dataset, which contains numerous categorical features and requires modelling complex interactions for accurate money laundering detection. The TabTransformer's success can be attributed to the effective Transformer architecture for such data characteristics (X. Huang et al., 2020), (Borisov et al., 2022). TabNet, which combines deep neural networks with decision trees, focuses on learning sparse and interpretable relationships within the data (Arik and Pfister, 2021), (Borisov et al., 2022). Its strength lies in feature selection, making it effective in datasets with many redundant or irrelevant features. Consequently, in scenarios where the dataset is smaller and feature relationships are predictive of outcomes, TabNet's effectiveness may be limited.



(a) Tab-AML models ROC curve.

(b) TabTransformer models ROC curve.

(c) Confusion matrix of the Tab-AML model with a true positive rate of 98% and false positive rate of 51%.

(d) Confusion matrix of the TabTransformer model with a true positive rate of 98% and false positive rate of 68%.

Figure 5.6: Comparative Analysis of ROC Curves and Confusion Matrices for the Tab-AML and TabTransformer Models, Highlighting the Performance at Equal True Positive Rates.

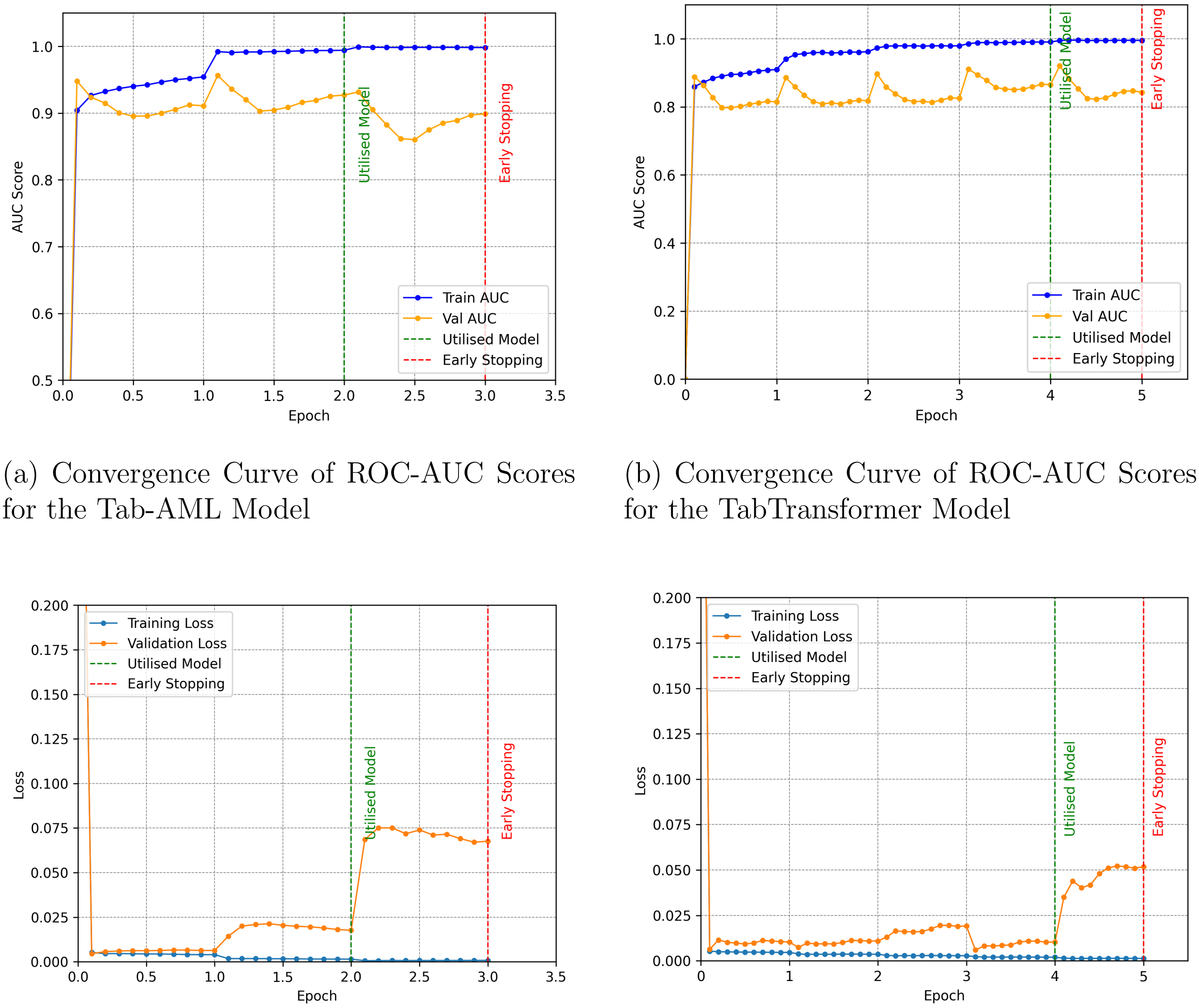
For this study's application, detecting money laundering transactions, the true positive rate is crucial to ensure that no suspicious activities are overlooked. Simultaneously, it is vital to minimise the false positive rate to reduce unnecessary costs and avoid the inefficient allocation of resources. I compared Tab-AML with the next best performing model which is the TabTransformer. I then looked at the ROC curve to see the model's performances across various thresholds and how that affects the true positive and false positive rates. After I decided on a threshold that produces a high true positive rate for both models and compare the confusion matrix. Figures 5.6 (a) and 5.6 (b) present the ROC curve of the Tab-AML and TabTransformer models. Tab-AML outperformed the TabTransformer model across all the thresholds. The TabTransformer's ROC curve (Figure 5.6 (b)) is flatter than the Tab-AML (Figure 5.6 (a)), hence producing more false positives as the number of truly positive transactions are detected. Figures 5.6 (c) and 5.6 (d) is the confusion matrix for the Tab-AML and TabTransformer models with thresholds set so both true positive rates are similar and high. Tab-AML attains a true positive rate of 98%, with a false positive rate of 51%. The TabTransformer model performs significantly worse, attaining a false positive rate of 68% and a similar true positive rate of 98%. Tab-AML reduces the false positive rate by 17% while maintaining the true positive rate compared to the best performing baseline model. Overall, the results demonstrate that the Tab-AML model has the ability to detect complex suspicious behaviours and patterns in the transactions that need to be monitored further.

In summary, although the TabTransformer architecture effectively manages complex, interaction-rich categorical data, the adapted architecture of Tab-AML, offers a superior capability to model complex patterns and long-range dependencies in transaction data, resulting in its greater performance.

5.7.1 Convergence Curve

In Figure 5.7, I present the convergence curves for maximising the ROC-AUC scores and minimising the loss, highlighting the learning dynamics of the Tab-AML and TabTransformer models. The Tab-AML model demonstrated faster convergence compared to the TabTransformer model. A sharp increase in both the training and validation ROC-AUC scores during the first epoch suggests a robust initial learning phase. The ROC-AUC scores for both models remain consistently similar, illustrating the models' ability to learn from the data without overfitting.

The key decision points during training are determined by the behaviour of the ROC-AUC scores. For the Tab-AML model, the ROC-AUC peak occurred at epoch two, while for the TabTransformer, the peak was at epoch four. Training was halted for both models one epoch after their peak due to the early stopping mechanism. The decision to stop training was informed by the absence of notable improvement in the validation datasets' ROC-AUC score beyond a threshold of $1e^{-4}$. This specific threshold was chosen after experimenting with alternatives ($1e^{-3}$ and $1e^{-5}$), as it provided the best balance between maximising model performance and effectively preventing overfitting. This approach was pivotal in preventing overfitting and optimising computational resources, ensuring that the models remained generalisable. Regarding the loss trends, after achieving a low initial loss in the first epoch, both models show varied patterns. The Tab-AML model's loss increased slightly in the second epoch and substantially in the third, aligning with the decrease in ROC-AUC scores and triggering early stopping. The TabTransformer model experienced a slight increase in loss up to the third epoch, a minor decrease in the fourth, and then a sharp increase in the fifth epoch. This increase coincided with the ROC-AUC scores peaking in the fourth epoch, leading to the early termination of



(a) Convergence Curve of ROC-AUC Scores for the Tab-AML Model

(b) Convergence Curve of ROC-AUC Scores for the TabTransformer Model

(c) Convergence Curve of Loss Scores for the Tab-AML Model

(d) Convergence Curve of Loss Scores for the TabTransformer Model

Figure 5.7: ROC-AUC and Loss Scores Convergence Curves for the TAB-AML and TabTransformer Models.

training and utilising the model from the fourth epoch.

Future research might benefit from exploring different thresholds or patience settings in the early stopping criterion to potentially extend model training where computationally feasible, aiming to enhance model robustness and performance.

5.8 Summary and Future Work

To conclude, this study contributes a transformer-based deep learning model to identify potential money laundering transactions. A shared embedding component is used to enhance the model's ability to learn various feature interactions, better detecting complex and subtle transaction patterns. The model incorporates a dual-masked transformer encoder structure, incorporating residual attention to enhance its performance. The first transformer block focuses on transaction flows, analysing interactions between sender and receiver account features at a micro-level, while the second transformer extends this analysis to a holistic examination of the complete feature set at a macro-level. This approach is complemented by the residual attention mechanism, to better detects subtle patterns in transactions. I also experimented with and evaluated the TabTransformer and TabNet models, neither of which had previously been utilised in the domain of money laundering detection, offering new perspectives in this field.

The SAML-D dataset was used to train, evaluate, and compare the models. The dataset went through standard machine learning pre-processing procedures, involving categorical encoding, log transformation, and standardisation. Hyperparameter tuning was conducted to optimise the various model's performances. All the models were initially compared based on the average of their ROC-AUC scores. This score was attained through conducting three individual experiments for each model, utilising a different seed value, to ensure robustness. The two best performing models were evaluated further by analysing their ROC curve and confusion matrices. To align the two model's capabilities, a specific threshold value was utilised, prioritising a high true positive rate. This approach was taken to minimise the occurrence of undetected suspicious transactions, which is critical in AML. The findings revealed that Tab-AML outperformed the other models, evidenced by an ROC-AUC score of 93.01%. Notably, the TabTransformer was the second best performing model in the study achieving an ROC-AUC score of 85.94%. Both transformer-based models surpassed the performance of XGBoost (81.12%), which is known for its exceptional ability in classification tasks within tabular data and in the AML literature. The deep learning model, TabNet, attained a ROC-AUC score of 80.61%. The Tab-AML model reduced the false positive rate by 17% without compromising the true positive rate (98%), when compared to the TabTransformer model. Achieving this type of balance is vital for AML.

Implementing a model like Tab-AML in financial institutions offers significant advantages, such as increased efficiency in detection capabilities while reducing operational costs. However, the deployment process presents several challenges, including the need for integration with pre-existing systems, access to high quality and complete datasets for training, and adherence to regulatory compliance. Financial institutions would require a proficient team to successfully incorporate such a model into their institution. Moreover, instead of replacing existing procedures, Tab-AML can be deployed as an additional tool on top of their existing rule-based methods, focusing on analysing alerted transactions that should be investigated further. This approach can enhance the existing AML framework by adding a layer of sophisticated analysis capable of uncovering complex patterns. It is important to note that privacy concerns are mitigated by the fact that the model is designed to be deployed within the secure infrastructure of financial institutions, using their own transactional data.

Our experiments were conducted on a synthetic dataset developed in collaboration with AML specialists. To further validate the models, I began work on a real dataset. Additional research on the application of transformer-based models could further build on this study's findings, deepening the understanding and enhancing the utilisation of these models within AML contexts. Further evaluation can be achieved by assessing the Tab-AML model on other synthetic datasets, attaining a more comprehensive understanding of the model's effectiveness and generalisability. Future work could also explore the use of feature engineering to potentially enhance the various model's performances. The explainability of the model, particularly the attention mechanism could be investigated further to offer a more in-depth understanding of its decision-making. Exploring Large Language Models (LLMs) with Tab-AML could further refine the investigation process by leveraging the model's explainability features, like feature importance scores. This strategy could enable investigators to examine alerts more effectively, facilitating quicker decisions on whether to file a SAR.

Overall, this study explores the use of transformers for money laundering detection and proposes a transformer-based model. The use of transformer encoders for money laundering detection is a largely unexplored research area for money laundering detection and this research can encourage future research into the application of transformer models in this domain.

Chapter 5 References

Akiba, Takuya et al. (2019). "Optuna: A Next-Generation Hyperparameter Optimization Framework". In: *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2623–2631.

Alarab, Ismail, Simant Prakoonwit, and Mohamed Ikbal Nacer (2020). "Competence of Graph Convolutional Networks for Anti-Money Laundering in Bitcoin Blockchain". In: *Proceedings of the 2020 5th International Conference on Machine Learning Technologies*. ICMLT '20. Beijing, China: Association for Computing Machinery, pp. 23–27. ISBN: 9781450377645. DOI: 10.1145/3409073.3409080.

Alshantti, Abdallah and Adil Rasheed (2021). "Self-Organising Map Based Framework for Investigating Accounts Suspected of Money Laundering". In: *Frontiers in Artificial Intelligence* 4, pp. 1–15. ISSN: 2624-8212. DOI: 10.3389/frai.2021.761925.

Alsuwailem, Abdullah A.S. and Akhtar Jamal Khan J. Saudagar (2020). "Anti-money laundering systems: a systematic literature review". In: *Journal of Money Laundering Control* 23.4, pp. 833–848. DOI: 10.1108/JMLC-02-2020-0018.

Arik, Sercan Ö. and Tomas Pfister (2021). "TabNet: Attentive Interpretable Tabular Learning". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 35.8, pp. 6679–6687. DOI: 10.1609/aaai.v35i8.16826.

Ba, Jimmy, Jamie Ryan Kiros, and Geoffrey E. Hinton (2016). *Layer Normalization*. arXiv preprint. Available from: https://arxiv.org/abs/1607.06450 [Accessed 17 Mar. 2024].

Bishop, Christopher M. and Nasser M. Nasrabadi (2006). *Pattern Recognition and Machine Learning*. Vol. 4. New York: Springer, p. 738.

Bisong, Ekaba (2019). "Logistic Regression". In: *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Berkeley, CA: Apress. DOI: 10.1007/978-1-4842-4470-8_20.

Borisov, Vadim et al. (2022). "Deep Neural Networks and Tabular Data: A Survey". In: *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–21. DOI: 10.1109/TNNLS.2022.3229161.

Bottou, Léon (2012). "Stochastic Gradient Descent Tricks". In: *Neural Networks: Tricks of the Trade*. Ed. by Grégoire Montavon, Genevieve B. Orr, and Klaus-Robert Müller. Vol. 7700. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-642-35289-8_25.

Breiman, Leo (2001). “Random Forests”. In: *Machine Learning* 45, pp. 5–32. DOI: 10.1023/A:1010933404324.

Chen, Tianqi and Carlos Guestrin (2016). “XGBoost: A Scalable Tree Boosting System”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ’16. San Francisco, California, USA: Association for Computing Machinery, pp. 785–794. ISBN: 9781450342322. DOI: 10.1145/2939672.2939785.

Dahouda, M. K. and I. Joe (2021). “A Deep-Learned Embedding Technique for Categorical Features Encoding”. In: *IEEE Access* 9, pp. 114381–114391. DOI: 10.1109/ACCESS.2021.3104357.

Domashova, Jenny and Natalia Mikhailina (2021). “Usage of machine learning methods for early detection of money laundering schemes”. In: *Procedia Computer Science* 190, pp. 184–192. ISSN: 1877-0509. DOI: 10.1016/j.procs.2021.06.033.

Fawcett, Tom (2006). “An introduction to ROC analysis”. In: *Pattern Recognition Letters* 27.8, pp. 861–874. ISSN: 0167-8655. DOI: 10.1016/j.patrec.2005.10.010.

Gal, Michal and Daniel L Rubinfeld (2019). “Data standardization”. In: *New York University Law Review* 94, pp. 737–770.

Granados, Oscar M. and Alejandro Vargas (2022). “The geometry of suspicious money laundering activities in financial networks”. In: *EPJ Data Science* 11.6. DOI: 10.1140/epjds/s13688-022-00318-w.

Guevara, Jorge, Olmer Garcia-Bedoya, and Oscar Granados (2020). “Machine Learning Methodologies Against Money Laundering in Non-Banking Correspondents”. In: *Applied Informatics*. Springer International Publishing, pp. 72–88. DOI: doi.org/10.1007/978-3-030-61702-8_6.

Hampo, J. A., E. C. Nwokorie, and J. N. Odii (2023). “A Web-Based KNN Money Laundering Detection System”. In: *European Journal of Theoretical and Applied Sciences* 1.4, pp. 277–288. DOI: 10.59324/ejtas.2023.1(4).27.

He, Ruining et al. (2021). *RealFormer: Transformer Likes Residual Attention*. arXiv preprint. Available from: https://doi.org/10.48550/arXiv.2012.11747 [Accessed 14 Mar. 2024].

Hendrycks, Dan and Kevin Gimpel (2016). *Gaussian Error Linear Units (GELUs)*. arXiv preprint. Available from: http://arxiv.org/abs/1606.08415 [Accessed 28 Mar. 2024].

Hervé, Taud and Mas Jean-François (2018). “Multilayer Perceptron (MLP)”. In: *Geomatic Approaches for Modeling Land Change Scenarios*. Cham: Springer, Chapter 27. DOI: 10.1007/978-3-319-60801-3_27.

Hoffer, Elad, Itay Hubara, and Daniel Soudry (2017). “Train longer, generalize better: closing the generalization gap in large batch training of neural networks”. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS’17. Long Beach, California, USA: Curran Associates Inc., pp. 1729–1739. ISBN: 9781510860964.

Huang, Xin et al. (2020). *TabTransformer: Tabular Data Modeling Using Contextual Embeddings*. arXiv preprint. Available from: https://doi.org/10.48550/arXiv.2012.06678 [Accessed 16 July. 2024].

Jullum, Martin et al. (2020). “Detecting money laundering transactions with machine learning”. In: *Journal of Money Laundering Control* 23.1, pp. 173–186. ISSN: 1368-5201. DOI: 10.1108/JMLC-07-2019-0055.

Ketenci, Utku Gorkem et al. (2021). “A Time-Frequency Based Suspicious Activity Detection for Anti-Money Laundering”. In: *IEEE Access* 9, pp. 59957–59967. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3072114.

Kute, Dattatray Vishnu et al. (2021). “Deep Learning and Explainable Artificial Intelligence Techniques Applied for Detecting Money Laundering–A Critical Review”. In: *IEEE Access* 9, pp. 82300–82317. DOI: 10.1109/ACCESS.2021.3086230.

Labib, Niveen M., Mai A. Rizka, and Ahmed E. M. Shokry (2020). “Survey of Machine Learning Approaches of Anti-money Laundering Techniques to Counter Terrorism Finance”. In: *Internet of Things—Applications and Future*. Ed. by Aziza Ghalwash et al. Vol. 114. Lecture Notes in Networks and Systems. Singapore: Springer. DOI: 10.1007/978-981-15-3075-3_5.

LeCun, Yann A. et al. (2012). “Efficient backprop”. In: *Neural Networks: Tricks of the Trade*. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 7700 LECTURE NO: Springer Verlag, pp. 9–48. DOI: 10.1007/978-3-642-35289-8_3.

Liashchynskyi, Petro B. and Pavlo Liashchynskyi (2019). *Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS*. arXiv preprint. Available from: https://arxiv.org/abs/1912.06059 [Accessed 21 Jan. 2024].

Liu, R. et al. (2011). “Research on anti-money laundering based on core decision tree algorithm”. In: *2011 Chinese Control and Decision Conference (CCDC)*. Mianyang, China, pp. 4322–4325. DOI: 10.1109/CCDC.2011.5968986.

Llugsi, R. et al. (2021). “Comparison between Adam, AdaMax and Adam W optimizers to implement a Weather Forecast based on Neural Networks for the Andean city of Quito”. In: *2021 IEEE Fifth Ecuador Technical Chapters Meeting (ETCM)*. Cuenca, Ecuador, pp. 1–6. DOI: 10.1109/ETCM53643.2021.9590681.

Lokanan, Mark E. (2022). “Predicting Money Laundering Using Machine Learning and Artificial Neural Networks Algorithms in Banks”. In: *Journal of Applied Security Research*. DOI: 10.1080/19361610.2022.2114744.

Luong, Thang, Hieu Pham, and Christopher D. Manning (2015). “Effective Approaches to Attention-based Neural Machine Translation”. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Ed. by Lluís Màrquez, Chris Callison-Burch, and Jian Su. Lisbon, Portugal: Association for Computational Linguistics, pp. 1412–1421. DOI: 10.18653/v1/D15-1166.

Lv, L. T., N. Ji, and J. L. Zhang (2008). “A RBF neural network model for anti-money laundering”. In: *International Conference on Wavelet Analysis and Pattern Recognition*. Vol. 1, pp. 209–215. ISBN: 2158-5709. DOI: 10.1109/ICWAPR.2008.4635778.

Mucherino, Antonio, Petraq J. Papajorgji, and Panos M. Pardalos (2009). “k-Nearest Neighbor Classification”. In: *Data Mining in Agriculture*. Vol. 34. Springer Optimization and Its Applications. New York, NY: Springer. DOI: 10.1007/978-0-387-88615-2_4.

Narkhede, M. V., P. P. Bartakke, and M. S. Sutaone (2022). “A review on weight initialization strategies for neural networks”. In: *Artificial Intelligence Review* 55, pp. 291–322. DOI: 10.1007/s10462-021-10033-z.

Oztas, B. et al. (2023a). “Enhancing Anti-Money Laundering: Development of a Synthetic Transaction Monitoring Dataset”. In: *2023 IEEE International Conference on e-Business*

Engineering (ICEBE). Sydney, Australia, pp. 47–54. DOI: 10.1109/ICEBE59045.2023.00028.

Oztas, Berkan et al. (2022). "Enhancing Transaction Monitoring Controls to Detect Money Laundering Using Machine Learning". In: *2022 IEEE International Conference on e-Business Engineering (ICEBE)*, pp. 26–28. DOI: 10.1109/ICEBE55470.2022.00014.

Paula, Ebberth L. et al. (2016). "Deep Learning Anomaly Detection as Support Fraud Investigation in Brazilian Exports and Anti-Money Laundering". In: *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 954–960. DOI: 10.1109/ICMLA.2016.0172.

Pedregosa, Fabian et al. (2018). *Scikit-learn: Machine Learning in Python.* arXiv: 1201.0490 [cs.LG].

Quinlan, J. R. (1986). "Induction of decision trees". In: *Machine Learning* 1, pp. 81–106. DOI: 10.1007/BF00116251.

Shokry, Amr Ehab Muhammed, Mohammed Abo Rizka, and Nevine Makram Labib (2020). "Counter terrorism finance by detecting money laundering hidden networks using unsupervised machine learning algorithm". In: *International Conference on e-Learning*, pp. 89–97. DOI: 10.33965/ict_csc_wbc_2020_2020081012.

Srivastava, Nitish et al. (2014). "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15.56, pp. 1929–1958.

Stojanović, Branka et al. (2021). "Follow the Trail: Machine Learning for Fraud Detection in Fintech Applications". In: *Sensors* 21.5. ISSN: 1424-8220. DOI: 10.3390/s21051594.

Sudjianto, Agus et al. (2010). "Statistical Methods for Fighting Financial Crimes". In: *Technometrics* 52.1, pp. 5–19. DOI: 10.1198/TECH.2010.07032.

Tanha, Jafar et al. (2020). "Boosting methods for multi-class imbalanced data classification: an experimental review". In: *Journal of Big Data* 7.70. DOI: 10.1186/s40537-020-00349-y.

Tatulli, Maria Paola et al. (2023). "HAMLET: A Transformer Based Approach for Money Laundering Detection". In: *Cyber Security, Cryptology, and Machine Learning: 7th International Symposium, CSCML 2023, Be'er Sheva, Israel, June 29–30, 2023, Proceedings.* Be'er Sheva, Israel: Springer-Verlag, pp. 234–250. ISBN: 978-3-031-34670-5. DOI: 10.1007/978-3-031-34671-2_17.

Tundis, A., S. Nemalikanti, and M. Mühlhäuser (2021). "Fighting organized crime by automatically detecting money laundering-related financial transactions". In: *The 16th International Conference on Availability, Reliability and Security.* Vol. 38, pp. 1–10. DOI: 10.1145/3465481.3469196.

Vaswani, Ashish et al. (2017). *Attention Is All You Need.* arXiv preprint. Available from: https://doi.org/10.48550/arXiv.1706.03762 [Accessed 21 May. 2024].

Zhang, Y. and P. Trubey (2019). "Machine Learning and Sampling Scheme: An Empirical Study of Money Laundering Detection". In: *Comput Economics* 54, pp. 1043–1063. DOI: 10.1007/s10614-018-9864-z.

Zhiyuan, Chen et al. (2021). "Variational Autoencoders and Wasserstein Generative Adversarial Networks for Improving the Anti-Money Laundering Process". In: *IEEE Access* 9, pp. 83762–83785. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3086359.

Chapter 6

Case Study: Evaluation Using a Real-World Dataset

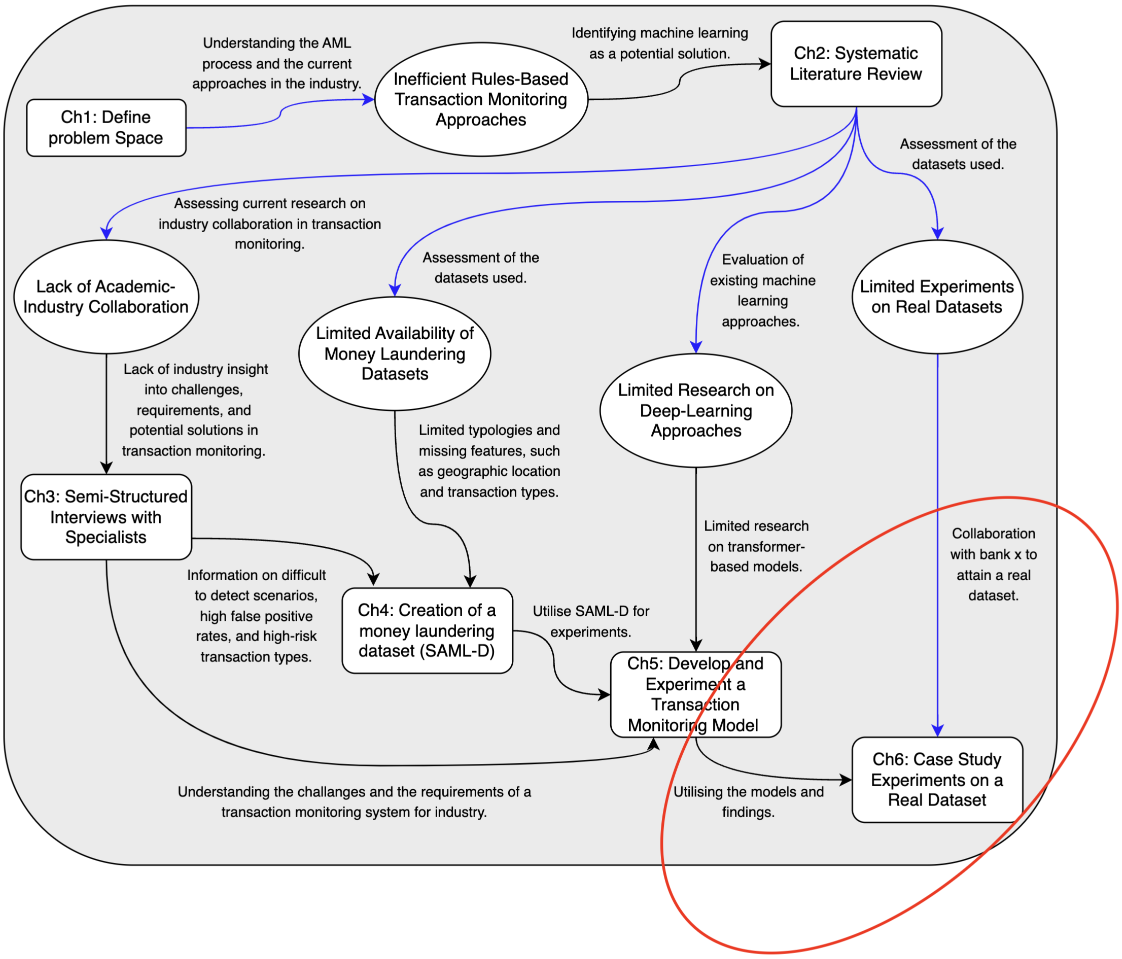

Ch1: Define problem Space
Understanding the AML process and the current approaches in the industry.
Inefficient Rules-Based Transaction Monitoring Approaches
Identifying machine learning as a potential solution.
Ch2: Systematic Literature Review
Assessment of the datasets used.
Assessing current research on industry collaboration in transaction monitoring.
Assessment of the datasets used.
Evaluation of existing machine learning approaches.
Lack of Academic-Industry Collaboration
Limited Availability of Money Laundering Datasets
Limited Research on Deep-Learning Approaches
Limited Experiments on Real Datasets
Lack of industry insight into challenges, requirements, and potential solutions in transaction monitoring.
Limited typologies and missing features, such as geographic location and transaction types.
Limited research on transformer-based models.
Collaboration with bank x to attain a real dataset.
Ch3: Semi-Structured Interviews with Specialists
Information on difficult to detect scenarios, high false positive rates, and high-risk transaction types.
Ch4: Creation of a money laundering dataset (SAML-D)
Utilise SAML-D for experiments.
Ch5: Develop and Experiment a Transaction Monitoring Model
Understanding the challanges and the requirements of a transaction monitoring system for industry.
Utilising the models and findings.
Ch6: Case Study Experiments on a Real Dataset

Figure 6.1: Overview of the PhD (extracted from Figure 1.3)

6.1 Introduction

In this chapter, I extend the analysis of the advanced deep learning models detailed in Chapter 5 by conducting experiments on a real transaction monitoring dataset, fulfilling objective (v). Figure 6.1 highlights the research motivation and the use of the models from the previous chapter. Furthermore, I evaluate the model's ability at a later task of the transaction monitoring process, beyond the initial monitoring of unprocessed transactions. A comparison of the model's performance against traditional machine learning methods that are prevalently utilised in the existing literature is conducted, with particular emphasis on XGBoost, which is widely acknowledged as one of the most effective models for transaction monitoring (Domashova and Mikhailina, 2021),(Berkan Oztas et al., 2022). The analysis is especially focused on models that incorporate transformers and attention mechanisms, namely Tab-AML, TabNet, and TabTransformer, to determine their efficacy on an authentic dataset. Since the same models detailed in Chapter 5 are used, a separate literature review is not included to avoid repetition. It is noteworthy that, to the best of my knowledge, such deep learning approaches have not yet been applied to this stage of transaction monitoring within financial institutions.

The dataset used in this chapter consists of grouped transactions aggregated to represent potential money laundering cases. This distinction is essential, as the models are predicting suspicious cases rather than individual transactions. Flagging a suspicious transaction typically involves analysing a single transaction in isolation or in relation to other linked transactions, focusing on characteristics such as large cash deposits, rapid transfers, or unusual cross-border payments. Identifying a suspicious case requires assessing an entity's aggregated transactions over time to identify patterns of illicit activity. This difference is crucial because the experiments in this chapter evaluate the models' ability to detect complex money laundering cases at a later stage in the transaction monitoring process.

The characteristics of the real dataset acquired from a collaborating bank are outlined, emphasising its distinctions from the synthetic datasets previously employed in this research and from traditional transaction monitoring datasets. Data preparation strategies and specific techniques implemented to optimise the data for analysis are then explained. Following these discussions, an overview of the machine learning and deep learning models being tested is presented. The experimental setup chapter then provides a detailed explanation of how the experiments were conducted. The results of these experiments were systematically presented and analysed. This analysis not only highlighted the comparative performance of the models but also shed light on the practical implications of deploying deep learning solutions in real-world transaction monitoring scenarios. Through this comprehensive examination, valuable insights were contributed to the field, informing future research directions.

6.2 The Dataset

The dataset utilised in this study was obtained through a collaboration with Bank X. To ensure confidentiality and adhere to the terms of the non-disclosure agreement (NDA), only limited information about the dataset can be disclosed. Nevertheless, a comprehensive understanding of its structure and characteristics is provided to highlight its distinctions from conventional transaction monitoring datasets.

The dataset provided by Bank X contains grouped transactions rather than individual transaction records. Unlike traditional raw transaction monitoring datasets, where each record represents a single transaction, this dataset has undergone a unique grouping process. Each row in the dataset represents a potential money laundering case involving multiple transactions, with features aggregating insightful information about the case. There are approximately 1.5 million cases created from aggregated transactions, with almost 10% of them labelled as suspicious. Whilst this datasets size is substantial, it is smaller than the synthetic dataset used previously. The proportion of suspicious cases is relatively high compared to traditional transaction monitoring datasets because the original transaction data has been processed to create and highlight cases with potential AML risks. The dataset is composed of around 90 features, with a 25-75% split between categorical and numerical features. All cases are labelled as either suspicious or non-suspicious, providing a clear target variable for the models. Before sharing the dataset, Bank X anonymised all personal information, such as names and addresses, to protect confidentiality.

6.2.1 Pre-Processing Methods

Given the sensitive nature of the dataset, features within the data cannot be disclosed. However, the data preparation steps are outlined to demonstrate the comprehensiveness of the pre-processing methods applied. Initially, each feature and its values were thoroughly analysed to identify any missing or invalid inputs to appropriately address them, ensuring that the final dataset was accurate, consistent, and well-prepared for model training. Notably, the handling of outliers was not deemed suitable in the processing pipeline due to the problem setting. The outstanding characteristics of a positive case are likely crucial to the learning phase. This approach ensures that significant unique data points are preserved and considered in the analysis.

To address missing values, each feature was carefully examined to determine the most appropriate imputation method. For numerical attributes, missing values were replaced using the mean of the attribute or zero, depending on the type of feature (Little and Rubin, 2019). This approach maintains the statistical properties of the feature and limits the loss of information by preventing the need to discard the row. For categorical attributes, missing values were imputed by introducing a separate ”Unknown” category (Little and Rubin, 2019). This method preserves the distribution of the data and prevents the loss of valuable information. In cases where missing values for a given feature could be inferred from other features, they were derived accordingly. Incorrectly entered values were replaced with zero or ”Unknown,” depending on the feature type, ensuring consistency, and preventing misleading interpretations.

Several features were dropped from the dataset due to duplication or redundancy. Removing such features ensured that the models would only be trained on relevant and unique information, enhancing their efficiency and interpretability (Cai et al., 2018). By eliminating redundant features, I reduced the dimensionality of the dataset and decreased computational complexity.

Categorical features were converted into numerical representations using appropriate encoding techniques. Label encoding or binary encoding was applied to categorical features, depending on the type of feature. This method provides a balance between preserving the distinctness of feature values and preventing the model from becoming overwhelmed by numerous categories (Hancock and Khoshgoftaar, 2020). Numerical features were standardised to ensure consistent scaling across different magnitudes and distributions. This step was crucial for models sensitive to feature scaling, such as neural networks and distance-based algorithms.

Standardisation helped align features on a comparable scale, preventing features with larger values from disproportionately influencing the models' performance (Dahouda and Joe, 2021).

By following these pre-processing steps, I ensured that the dataset was clean, consistent, and optimised for analysis using both deep learning models and traditional machine learning methods. The rigorous handling of missing data, irrelevant features, standardisation, and encoding ensured that the models would learn effectively from the data.

6.3 Experiment Set-up

This research aims to further evaluate the effectiveness of the deep learning model and baseline models in transaction monitoring using a real dataset. The approach includes refining the models through hyperparameter optimisation on the training and validation datasets. Once the optimal parameters are identified, they are implemented on a test dataset to evaluate the models' ability to detect suspicious cases. The peak performance outcomes of each model, with the chosen parameters on both the validation and test datasets, are presented in Tables 6.1 and 6.2.

Our experimental design adheres to an 80-10-10 time based split for the training, validation, and testing phases. Initially, all cases were organised in chronological order by their date of creation. The training data consists of 80% (1,129,471 cases) of the dataset used to fit the models to the data characteristics. The remaining data is evenly divided between the validation and test datasets, each comprising 10% of the cases (141,183 cases). The validation set is used for refining the hyperparameters, and the test set for assessing the models' performance. I carefully monitored for overfitting by analysing performance on the validation and test datasets, ensuring the models did not learn the training data to an extent that would impair their performance on new data.

6.3.1 Evaluation Metric

To assess the effectiveness of various models, I chose the ROC-AUC score as the evaluation metric (Ling, J. Huang, and H. Zhang, 2003). By averaging the ROC scores from experiments conducted over three different seed values, I aimed to generalise the results. The objective is to evaluate the models' ability to distinguish between cases of money laundering and normal financial activities. Solely using accuracy as a metric would not capture the complexity required for this analysis, due to data imbalance.

Further analysis was done by plotting the model's ROC curves and creating confusion matrices at chosen thresholds. The ROC curve, which plots the true positive rate against the false positive rate at various decision thresholds, visually represents the model's performance. The area beneath this curve, referred to as the ROC-AUC score, quantifies a model's capacity to discriminate between classes. This metric ranges from 0.5, where the model performs equally to randomly guessing, to 1, indicating perfect classification. The confusion matrices present the model's performance in predicting positive and negative cases, showing both correct and incorrect classifications (Canbek et al., 2017). These evaluation metrics, frequently used in the AML domain, provide essential insights into the models' capabilities.

6.3.2 The Models

In this study, I employ advanced deep learning models that utilise transformers and attention mechanisms, specifically TabTransformer, TabAML, and TabNet (utilised in Chapter 5). These models are relevant to the AML domain due to their ability to handle complex and high-dimensional tabular data, a common characteristic of financial transaction datasets. In addition to the advanced deep learning models, I employ several traditional machine learning models to compare their performance in the AML context. These baseline models are widely used in the literature and provide a benchmark.

TabTransformer (X. Huang et al., 2020) is a deep tabular data modelling architecture for supervised and semi-supervised learning. It leverages attention mechanisms to capture dependencies and interactions between different features, making it exceptional for tasks such as transaction monitoring. The model starts by transforming categorical features into dense embeddings, ensuring that each category is represented as a continuous vector. At its core, transformer encoder blocks apply multi-head self-attention and feed-forward layers to capture interactions between features. This self-attention mechanism allows the model to dynamically adjust the importance of each feature based on the input data, enhancing its final prediction. In this case, the prediction is the classification of AML cases as suspicious or not.

Tab-AML is a transformer-based model created for transaction monitoring. The primary objective of Tab-AML is to accurately predict and identify suspicious transactions that require further investigation and potential reporting to authorities. Tab-AML achieves this by leveraging transformers encoders to handle complex categorical features in tabular datasets. Extensive details on the Tab-AML architecture can be found in Chapter 5. The initial stage of the Tab-AML architecture involves embedding categorical features, preparing continuous features, and utilising a shared embedding mechanism. Tab-AML incorporates two transformer encoders with residual attention mechanisms. The first transformer analyses transaction flows by examining the interactions between sender and receiver accounts, crucial for identifying suspicious patterns and money movements. The second transformer builds upon these insights by examining the complete transaction features, including additional details. This dual-layered masked approach ensures a thorough analysis, however, due to the dataset provided by Bank X, which is aggregated transactions into cases that do not link with each other, I have modified Tab-AML, so it only uses a singular encoder layer with residual attention to analyse the complete feature set.

The final deep learning model I use is TabNet (Arik and Pfister, 2021), designed for tabular data and aims to combine interpretability with the learning capabilities of deep learning. TabNet uses sequential attention to process data in a manner that mimics decision paths. The model processes features through feature transformer blocks, converting features into a more informative representation. The transformed features are then passed through transformers, to select the most relevant features for subsequent layers. This mechanism ensures that only the relevant features are considered at each step, resembling the decision paths in a decision tree. Outputs of the attentive transformers are aggregated across multiple decision steps, enhancing the models' predictive accuracy.

I employed various baseline models regularly used in the AML literature to evaluate and compare their performance in identifying suspicious cases compared to the deep learning models. These models include XGBoost, known for its high performance and efficiency on tabular data and within the AML domain. Logistic Regression was included for its simple yet powerful method for binary classification and offering high interpretability. I also used KNN, an instance-based learning algorithm that classifies samples based on the majority class among their k-nearest neighbours. Decision Trees was also included, specifically the CART model, a hierarchical model that splits data into subsets based on feature values, capturing non-linear relationships. Random Forests, an ensemble learning method that builds multiple decision trees and merges their outputs to improve classification, was included for its robustness and performance with large and high-dimensional datasets.

6.3.3 Hyperparameter Tuning

This section describes the method for adjusting hyperparameters, detailing the range of values tested for each parameter with the selected settings outlined in Tables 6.1 and 6.2. Hyperparameters not explicitly tuned were set to their default values. Due to computational constraints, I opted for manual hyperparameter tuning for the deep learning models rather than employing computationally intensive techniques such as GridSearch or Optuna. Specifically, I adjusted each hyperparameter individually, systematically varying one parameter at a time while holding all others constant to clearly isolate and evaluate its impact on model performance. For the baseline models, however, GridSearchCV was utilised for parameter refinement, as this approach was computationally feasible.

Our tuning efforts were primarily aimed at maximising the ROC-AUC score on the validation dataset, ensuring that enhancements in model performance would translate into greater generalisability. Despite the inherent limitations of manual tuning, such as the intensive time requirement and the potential for not exploring all hyperparameter dimensions fully (Shwartz-Ziv and Armon, 2022), this method yielded insights into the sensitivity of the models to various hyperparameter changes. With more computational resources, future research could explore broader optimisation strategies to improve the hyperparameter tuning process.

Deep Learning Models

In this study, I fine-tuned the performance of the Tab-AML and TabTransformer models by experimenting with various hyperparameters. I implemented Binary Cross-Entropy with Logits Loss (BCEwithLogitLoss) for both models, chosen for its advantageous numerical stability. The learning rate was configured to $1e^{-4}$ for the Tab-AML model and $1e^{-3}$ for the TabTransformer model. Both models utilised the AdamW optimiser, using its default parameters. Additionally, it is important to note that in the experiments, the number of layers was consistently set to half the number of heads.

For the TabNet model, I utilised the guidelines provided in its original publication and employed a PyTorch-based implementation. I used cross-entropy loss as the loss function and similarly used the Adam optimiser with its standard settings. The learning rate was set to $3e^{-2}$ with a scheduler that lowered the learning rate to 90% of the current learning rate every 30 steps. The deep learning models and the specific hyperparameters I explored are detailed below:

1. **TabTransformer** was tested using various parameter settings. The dimension was tested with values in $d \in \{16, 32, 64, 128\}$, the batch size was tested with values in $B \in \{256, 512, 1024, 2048\}$, the number of heads was tested with $Heads \in \{4, 8, 16, 32, 64\}$, the number of layers was tested with $Layers \in \{2, 4, 8, 16, 32\}$, learning rates were tested with values in $lr \in \{1e^{-5}, 1e^{-4}, 1e^{-3}, 1e^{-2}\}$, and attention dropout and feed forward dropout were tested with rates in $\{0\%, 10\%, 20\%\}$.

2. **TabAML** was evaluated with a range of parameter settings. The dimension was tested with values in $d \in \{16, 32, 64, 128\}$, the batch size was tested with values in $B \in \{256, 512, 1024, 2048\}$, the number of heads was tested with $Heads \in \{4, 8, 16, 32, 64\}$, the number of layers was tested with $Layers \in \{2, 4, 8, 16, 32\}$, learning rates were tested with values in $lr \in \{1e^{-5}, 1e^{-4}, 1e^{-3}, 1e^{-2}\}$, and residual and MLP dropout rates were tested with values in $\{0\%, 10\%, 20\%\}$.

3. **TabNet** was explored using various parameter values. The width of the decision prediction layer and attention embedding for each mask were tested with values in $N_d = N_a \in \{16, 32, 64, 128\}$, the number of decision steps was tested with values in $N_{\text{steps}} \in \{3, 5, 7\}$, the feature re-usage in the masks was tested with values in $\gamma \in \{1, 1.5, 2\}$, the momentum was tested with values in $M \in \{0.01, 0.1, 1, 10, 100\}$, the batch size was tested with values in $B \in \{4096, 8192, 16384, 32768\}$, and the virtual batch size was tested with values in $VB \in \{512, 1024, 2048, 4096\}$.

Baseline Models

Each baseline model was optimised using GridSearchCV, evaluating their effectiveness through the ROC-AUC score. This strategy was crucial for both assessing the models and enhancing the comprehension of the machine learning techniques employed to detect money laundering. The implementations were carried out using the Scikit-learn and XGBoost libraries. Details of the chosen baseline methods are presented below:

1. **Logistic Regression** (Bisong, 2019), is tested using the lbfgs solver with L2 Regularisation Strength values of $C \in \{0.01, 0.1, 1, 10, 100\}$.

2. **K-Nearest Neighbours (KNN)** (Mucherino, Papajorgji, and Pardalos, 2009), is evaluated with a range of neighbours: $N \in \{60, 70, 80, 90\}$.

3. **Decision Tree,(CART)** (Quinlan, 1986), are explored for Maximum Tree Depth values of $D \in \{2, 4, 8, 16\}$ and Minimum Samples Split, $S \in \{4, 8, 12, 16\}$.

4. **Random Forest** (Breiman, 2001), is tested for Maximum Tree Depth values of $D \in \{2, 4, 8, 12\}$ and Maximum Features to Consider per Split, $F \in \{2, 4, 8, 12\}$.

5. **XGBoost** (T. Chen and Guestrin, 2016), is evaluated for Maximum Tree Depth, $D \in \{2, 4, 8, 16\}$, Number of gradient boosted trees, $E \in \{100, 200, 300\}$, and Learning Rate options, $L \in \{0.1, 0.2, 0.3\}$.

6. **Gaussian Naive Bayes (GaussianNB)** (Bishop and Nasrabadi, 2006), is also included to benchmark its effectiveness without hyperparameter tuning.

Models	**ROC-AUC Score**	**Hyperparameters**
Tab-AML	84.29 (0.09)	O = AdamW, $lr = 1e^{-4}$, $d = 64$, $B = 1024$, Heads = 16, Layers = 8
TabTransformer	84.44 (0.05)	O = AdamW, $lr = 1e^{-3}$, $d = 64$, $B = 512$, Heads = 16, Layers = 8
TabNet	85.18 (0.06)	O = Adam, $lr = 3e^{-2}$, $N_d = N_a = 64$, $N_{\text{steps}} = 3$, $\gamma = 1$, $M = 0.2$, $B = 16384$, $VB = 512$
Baseline Models		
Logistic Regression	75.52	$C = 100$
K-Nearest Neighbours	81.81	$N = 11$
Decision Tree, (CART)	82.43	$D = 8$, $S = 12$
Random Forest	87.18	$D = 32$, $F = 4$
XGBoost	**87.61**	$D = 12$, $E = 200$, $L = 0.1$,
Gaussian Naive Bayes	52.71	-

Table 6.1: ROC-AUC Scores and Best Hyperparameters for the Models on the Validation Dataset. Displays the average ROC-AUC scores, computed over three different seeds, alongside the optimal hyperparameters identified for each model.

Models	**ROC-AUC Score**
Tab-AML	84.89 (0.15)
TabTransformer	85.10 (0.31)
TabNet	86.10 (0.13)
Baseline Models	
Logistic Regression	77.70
K-Nearest Neighbours	82.23
Decision Tree, (CART)	83.11
Random Forest	87.39
XGBoost	**88.73**
Gaussian Naive Bayes	51.62

Table 6.2: ROC-AUC Scores of the Models on the Test Dataset. Displays the average ROC-AUC scores for each model, computed over three different seeds.

6.4 Results and Discussion

To evaluate the models with the optimised parameters, I conducted experiments on the test dataset. Tables 6.1 and 6.2 present the ROC-AUC Scores on the validation and test datasets. XGBoost performed the best, attaining an ROC-AUC score of 87.61% on the validation dataset and 88.73% on the test dataset, outperforming the deep learning models. This indicates that XGBoost is highly effective in classifying suspicious cases, possibly due to its ability to handle complex feature interactions, capture nonlinear relationships, and effectively manage imbalanced datasets through techniques like regularisation and optimised tree boosting. TabNet slightly outperformed the other deep learning models with an ROC-AUC score of 85.18% on the validation dataset and 86.1% on the test dataset. This model's performance showcases its potential in handling tabular data by leveraging its sequential attention mechanism, which can capture complex feature interactions. The Tab-AML and TabTransformer models achieved similar ROC-AUC scores, despite the inclusion of the residual attention mechanism for the Tab-AML model. This suggests that while the residual attention mechanism is beneficial for certain datasets it might not be for this specific dataset and task. Regarding the baseline models, Random Forests was the second-best performing model, followed by Decision Trees. Logistic Regression and Gaussian Naive Bayes produced significantly lower results compared to the rest of the models. Interestingly, the results obtained on the test dataset are considerably better than on the validation dataset. I believe this could be because I did a time-based split and Bank X continually worked on its data quality throughout the data collection period. As a result, the more recent data in the test set might be cleaner and more representative of the true distribution, leading to better model performance.

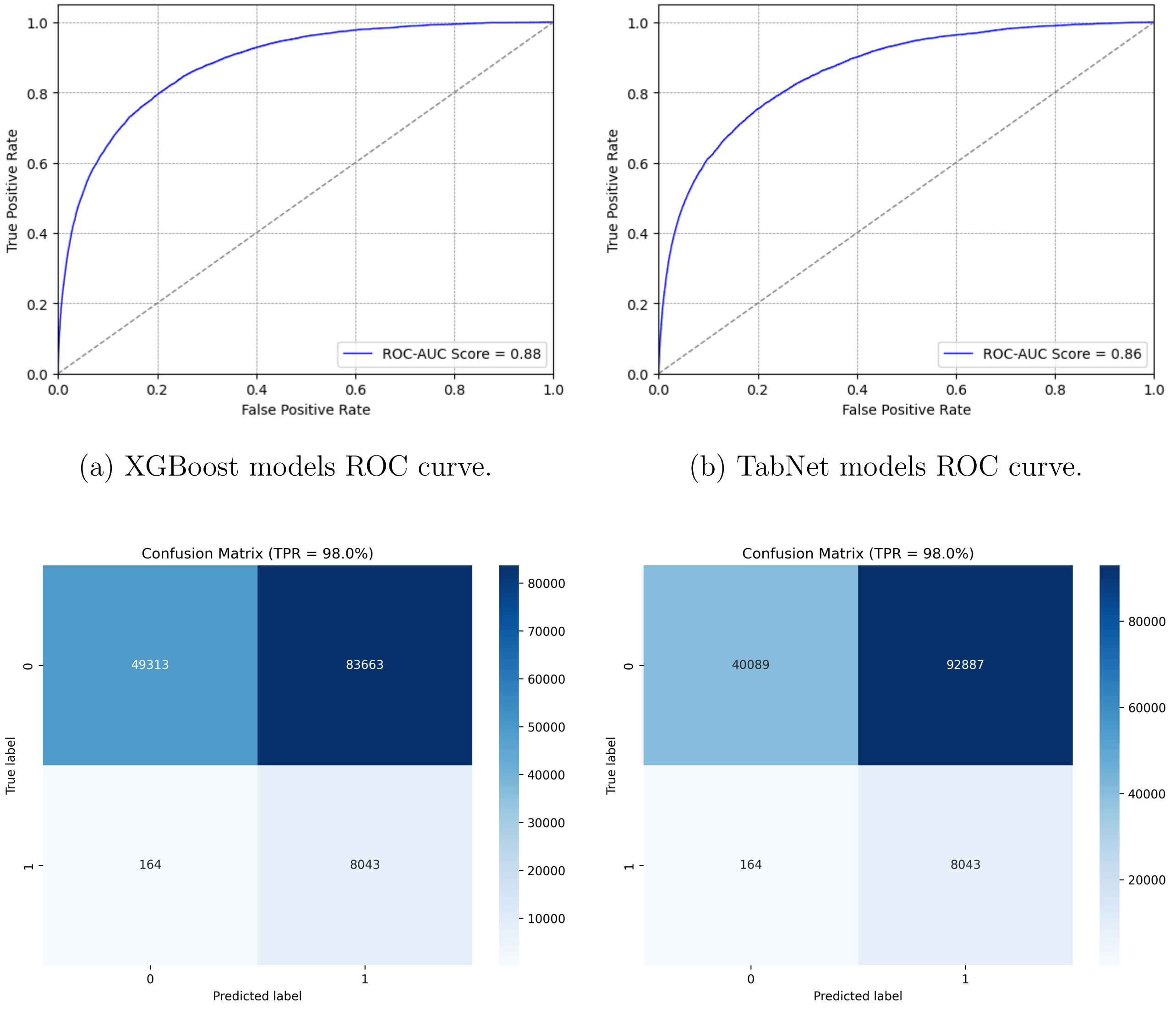
I compared XGBoost and TabNet as they were the best performing baseline and deep learning models. Figure 6.2 displays the ROC-AUC curves of the models along with the confusion matrix. The threshold for the confusion matrix is set so that the true positive rate is 98%, which is crucial for financial institutions. Ensuring a high true positive rate is essential for minimising the risk of missing fraudulent transactions, which can have severe financial implications (Kute et al., 2021). XGBoost attained a false positive rate of 62.92%, whereas the TabNet model got 69.86%. The XGBoost model reduces the false positive rate by 6.94% while retaining the same true positive rate. These results are both improvements on rule-based methods that are known to generate higher false positive rates (Vorobyev and Krivitskaya, 2022). This reduction in false positives is significant for practical applications, as it implies fewer unnecessary investigations and better resource allocation.

Overall, the results underscore the effectiveness of tree-based models, particularly XGBoost, in classifying AML cases. Despite the strong performance of deep learning models like TabNet, the simpler and more efficient tree-based models remain superior, especially when the goal is to balance predictive performance with practical deployment considerations.

6.4.1 Findings in Context with Previous Results

The primary distinction between the SAML-D (B. Oztas et al., 2023a) and Bank X's datasets lies in their composition and the stages they represent in the transaction monitoring process. The SAML-D dataset focuses on unprocessed transactions, representing the first step in identifying potential fraud. It consists of individual transactions over time, with each transaction exhibiting behaviours and structures that might indicate suspicious activities. SAML-D emphasises the importance of both the features of individual transactions and their interconnections, which are crucial for tracing money flows and identifying patterns indicative of illicit financial activities (B. Oztas et al., 2023a). The real dataset from Bank X represents a later stage in the process. It is case-oriented, created by first reducing the number of transactions to those with higher suspicion and then aggregating connected transactions into comprehensive cases. Each case focuses on aggregated features rather than individual transactions. This aggregation process results in minimal direct connections between different cases, making the dataset less effective for detecting patterns that span across multiple transactions, unlike the SAML-D dataset. Hence, while both datasets aim to monitor transactions for suspicious activities, they cater to different aspects of the transaction monitoring process.

The experiments highlighted that the best-performing models were closely aligned with the unique characteristics of each dataset. For the SAML-D dataset, transformer-based models such as Tab-AML and TabTransformer demonstrated superior performance. This dataset, due to its characteristics and purpose, necessitated models that could adeptly manage complex time-based relationships and patterns. The architecture of transformer models, equipped with attention mechanisms, allows them to emphasise different parts of the data dynamically, making them especially effective for tabular data that contains intricate dependencies (Vaswani et al., 2017). These models excel at detecting complex transaction flows, crucial for identify-



(a) XGBoost models ROC curve.

(b) TabNet models ROC curve.

(c) Confusion matrix of the XGBoost model with a true positive rate of 98% and false positive rate of 51%.

(d) Confusion matrix of the TabNet model with a true positive rate of 98% and false positive rate of 68%.

Figure 6.2: Comparative Analysis of ROC Curves and Confusion Matrices for the XGBoost and TabNet Models, Highlighting the Performance at Equal True Positive Rates.

ing suspicious activities (Starnini et al., 2021). The real dataset from Bank X, which is more case-based and aggregated information, XGBoost emerged as the most effective model. This dataset's structure, where cases are summarised with feature sets rather than detailed transactional relationships requires a robust model in feature handling. XGBoost, with its gradient boosting framework, leverages an ensemble of decision trees that can independently capture various aspects of the data. This model's ability to manage a wide range of data types and its computational efficiency makes it ideal for identifying suspicious cases. The ability to handle diverse and comprehensive features effectively, without the need for explicit inter-case linkages, highlights the reason XGBoost was the standout model for Bank X's dataset (T. Chen and Guestrin, 2016). When comparing the deep learning models to XGBoost, they showed slightly reduced but comparable performance.

Dataset	**Model**	**ROC-AUC**	**TP**	**TN**	**FP**	**FN**
SAML-D	Tab-AML	**93.01**	1085	461186	488197	18
	TabTransformer	85.94	1084	303597	645786	19
Real Dataset	Tab-AML	84.89	8044	37433	95543	163
	TabTransformer	**85.10**	8042	38813	94163	165

Table 6.3: Performance Metrics Comparison of Tab-AML and TabTransformer on the two datasets.

In the assessment of the transformer-based models on the two datasets, a notable point of difference was the effect of the residual attention. Residual attention enhances the standard attention mechanism by capturing long-range dependencies more effectively. It achieves this by integrating the attention scores from previous layers into the current layer's attention computation before applying the Softmax function (R. He et al., 2021). Despite these advanced capabilities, the experiments revealed that residual attention did not enhance performance on Bank X's datasets, which lack inter-row connections. However, for the SAML-D dataset, where the relationships between transactions are critical, the residual attention mechanism played a pivotal role in achieving a deeper understanding of the data. This was evident in the performances of Tab-AML and TabTransformer on the two datasets, presented in Table 6.3. On the SAML-D dataset, Tab-AML outperformed TabTransformer, attaining an ROC-AUC score of 93.01%, reducing the false positive rate by 17% compared to TabTransformer. On Bank X's dataset, TabTransformer slightly outperformed Tab-AML, attaining a ROC-AUC score of 86.10%, and reduced the false positive rate by 1.04% more than Tab-AML. In comparison, their performance on Bank X's dataset was less distinguished, highlighting the value of residual attention in contexts where inter-transactional relationships are essential for effective analysis.

In summary, the effectiveness of different models on the SAML-D and Bank X datasets illustrates the importance of aligning model capabilities with the specific task, characteristics, and requirements of the data being analysed. Transformer-based models, particularly those utilising shared embeddings and residual attention, excel in environments characterised by intricate interconnections and complex patterns. This makes them well-suited for the initial stage of transaction monitoring, where they can effectively process and analyse unstructured or unprocessed transaction data. XGBoost's ability to manage diverse and comprehensive features independently makes it better suited for processed transactions aggregated into case structures.

Understanding these distinctions is crucial for effectively applying machine learning models to specialised tasks and data types within transaction monitoring.

6.4.2 Further Analysis on Features

This section utilises the best performing model, XGBoost, to conduct a detailed examination of the feature set and assess the model's performance with selected subsets of features. This research was motivated by the desire to identify the most crucial features that enable the model to detect money laundering within the dataset, aiming to enhance explainability and understanding. Additionally, I sought to determine whether better performance could be achieved by using a subset of features, thereby reducing the influence of less relevant variables and computational costs. However, due to the NDA, I was unable to discuss specific features directly, so they were grouped into categories for analysis and insights. Hyperparameter tuning was performed using GridSearchCV for all experiments, employing the values presented previously in Section 6.3.3 on Hyperparameter tuning. To ensure confidentiality and adhere to the terms of the NDA, only limited information about the dataset can be disclosed.

Initially, the features were grouped into four categories: customer profile, transaction behaviour, risk indicators, and country/region. Each group was tested in isolation to determine its individual impact on the model's performance, presented in Figure 6.3. The results indicated that the risk identification features, followed by transaction behaviour features, were the most influential groups. The customer profile features ranked third, while the country/region features demonstrated the lowest performance, suggesting that they are not particularly effective in identifying suspicious cases when used in isolation. The greater performance of the risk indicator features can be attributed to their inclusion of historical outputs related to the client's prior activities. These features serve as indicators of past behaviour patterns that may signify risk, effectively creating a feedback loop from previous outcomes. This observation aligns with the concept that including historical outcomes to detect money laundering is crucial for a well-designed transaction monitoring approach (B. Oztas et al., 2023b). The transaction behaviour features also demonstrated strong performance, as they enable the detection of patterns that deviate from a client's typical behaviour, providing indications of potential suspicious activity. The customer profile features, while informative, is not as indicative of fraudulent activity since they often consist of static demographic information.

To optimise computational resources, which is crucial when monitoring millions of transactions daily, I experimented with using fewer features while aiming to maintain or improve model performance. I began by experimenting on the complete feature set and employed SHAP (SHapley Additive exPlanations) to identify the most important features for detecting suspicious cases. SHAP is a model-agnostic interpretability method based on Shapley values from cooperative game theory (Lundberg and Lee, 2017). It explains individual predictions by computing the contribution of each feature to the prediction, considering all possible combinations of features. This approach provides a measure of feature importance, allowing for a detailed understanding of the model's decision-making. Compared to other feature importance methods such as permutation importance (Altmann et al., 2010) or mean decrease impurity (Louppe et al., 2013), SHAP offers the advantage of accounting for feature interaction effects.

As presented in Figure 6.4, using the top 30 features onward yields results comparable to

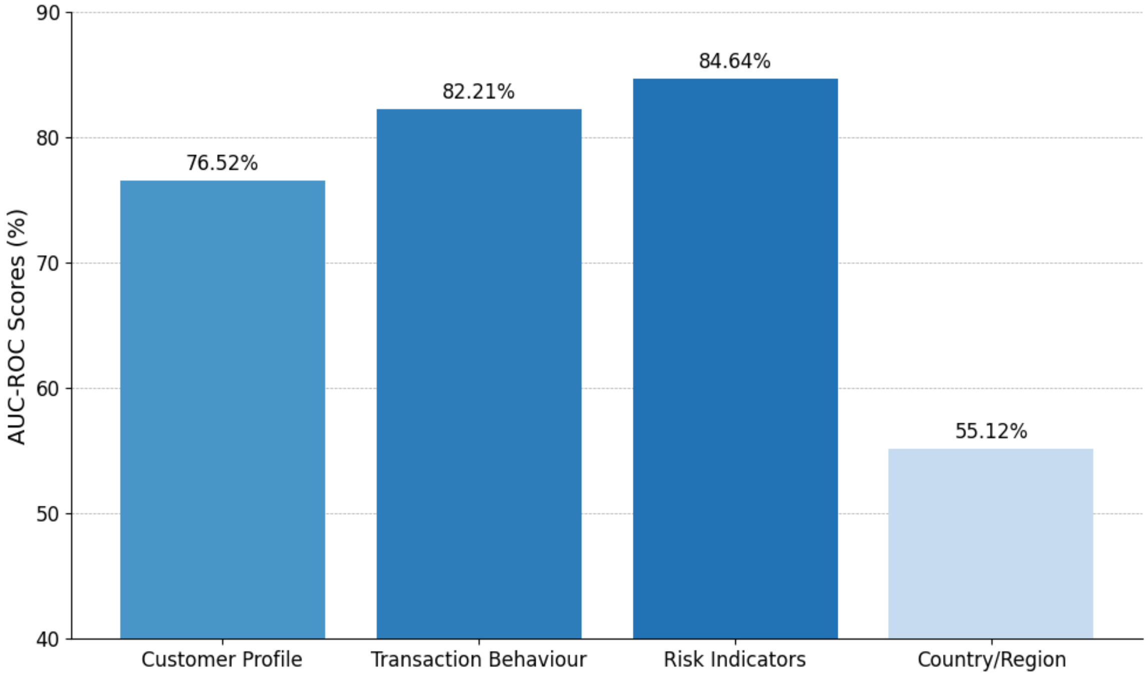


Figure 6.3: ROC-AUC Scores from XGBoost Models Utilising Grouped Features: This figure illustrates the performance impact of categorically grouped features on the model's predictive accuracy, highlighting the differential effectiveness of each feature group in isolation.

those obtained with the complete feature set. Models utilising the top 30, 40, and 50 SHAP-selected features slightly outperformed the model using all features, with the top 40 features achieving the best performance. Specifically, the top 40 features resulted in a 0.12% improvement in the ROC-AUC score compared to the complete dataset. This marginal improvement suggests that the additional features beyond the top 40 may introduce noise or redundant information, not providing any additional value to the model. By reducing the feature set to the most impacful features, I enhanceed computational efficiency and slightly improve model performance. In practical applications reducing computational complexity is essential. Fewer features lead to faster processing times, enabling quicker detection of fraudulent activities, which is critical in mitigating financial risks.

Further analysis was conducted to examine the distribution of feature groups among the top 40 SHAP-selected features. Figure 6.5 presents the percentage of each group within these top features. Although risk identification features were the most influential when used in isolation, transaction behaviour features constituted the majority (60%) of the top 40 features. This is due to the larger number of transaction behaviour features available compared to other groups. As anticipated, the country/region features were included the least among the top features, confirming their limited individual predictive power in this context.

6.5 Summary

In conclusion, this chapter focuses on enhancing the analysis of the advanced deep learning models, that incorporate transformers and attention mechanisms, using a real transaction monitoring dataset. The aim is to evaluate these models' effectiveness beyond the initial phase of

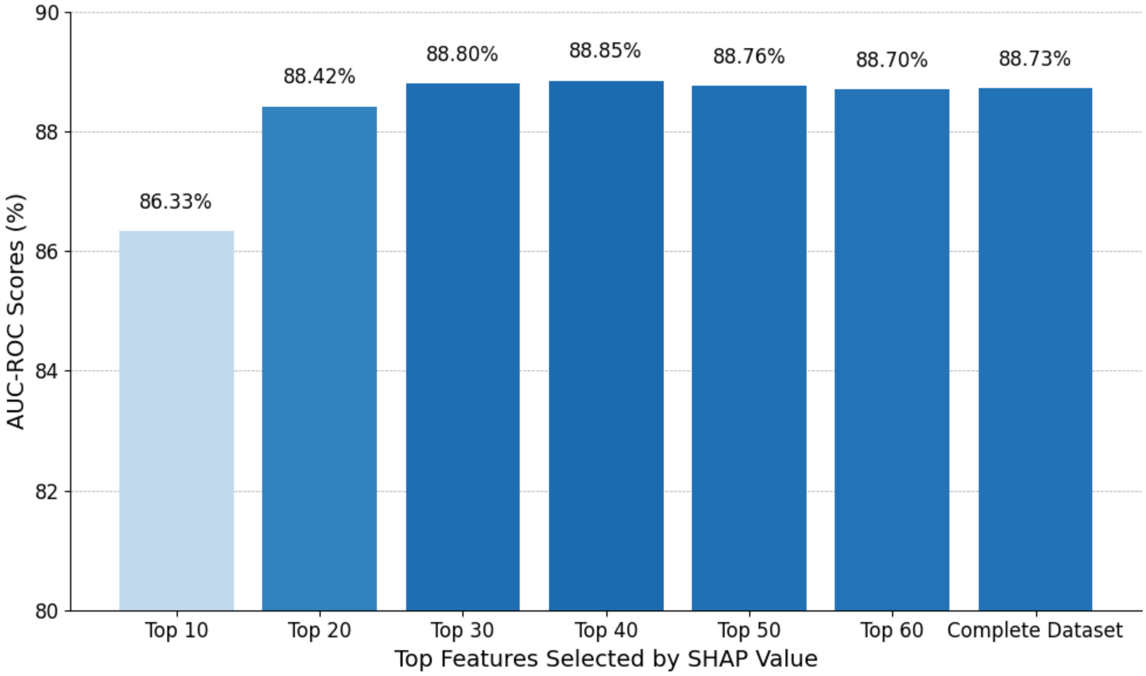


Figure 6.4: ROC-AUC Scores from XGBoost Using SHAP-Value Ranked Top Features: This graph displays the model's performance as more top-ranked SHAP features are included.

monitoring unprocessed transactions and to compare them against traditional machine learning methods, particularly XGBoost, known for its efficacy in transaction monitoring. This analysis includes an exploration of the real dataset provided by a collaborating bank, highlighting its distinctive characteristics in comparison to both synthetic and conventional datasets. The data preparation and optimisation techniques employed for the experiments are detailed. The experimental results are presented and analysed, demonstrating the performance of these models relative to traditional methods, but also the practical implications of deploying deep learning models for specific tasks in transaction monitoring. Additionally, the chapter conducted a detailed examination of the feature set and assessed the best model's performance with selected subsets of features. This comprehensive evaluation contributes valuable insights to the field, guiding future research directions.

The findings of this research distinguished XGBoost as the most effective model on the dataset provided by Bank X, achieving an ROC-AUC score of 87.61% in the validation set and 88.73% in the test set. TabNet proved to be the best among the deep learning models, leading with an ROC-AUC score of 85.18% in the validation dataset and 86.10% in the test dataset. The Tab-AML and TabTransformer models demonstrated comparable levels of performance. Notably, the Tab-AML model incorporates a residual attention mechanism, which was identified to be advantageous where inter-transactional relationships are crucial for effective detection. The analysis highlights the critical role of matching the model's capabilities with the unique features and demands of the specific task and dataset. Specifically, transformer-based models are particularly proficient in identifying complex patterns in unprocessed transactions, whereas XGBoost excels with processed data grouped into case structures. These distinctions are essential for the effective implementation of machine learning approaches in targeted transaction monitoring activities. Also, the study found that using the top 40 SHAP-selected features in the XGBoost model improved the ROC-AUC score by 0.12%. This enhancement indicates that

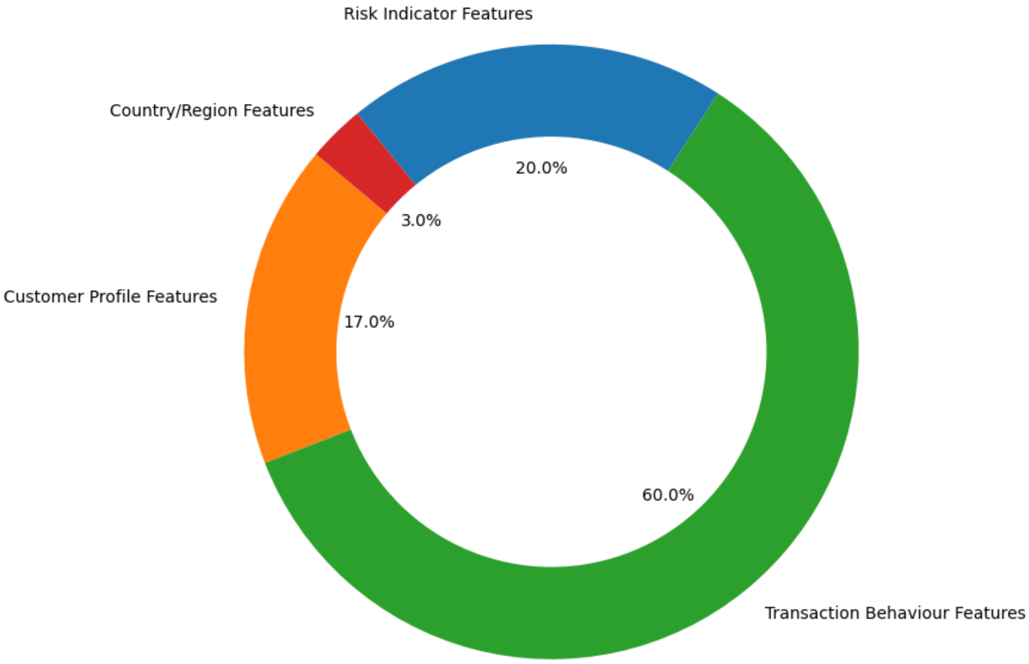


Figure 6.5: Distribution of Feature Group Among the Top 40 SHAP-Selected Features: The pie quantifying the contribution of each group to the model's predictive power as selected by SHAP values.

reducing the feature set enhances model performance and computational efficiency, crucial for effective transaction monitoring. The importance of feedback data was particularly highlighted in the performance of the 'Risk Indicator' features.

Despite the valuable insights gained from the study, it does have limitations. The significant computational demands of hyperparameter tuning in deep learning models are a notable constraint. Future work could investigate a more comprehensive hyperparameter tuning strategy, such as using Optuna. Moreover, the experiments were limited to a single real-world dataset, so further validation across more datasets could be beneficial to generalise the findings. Due to privacy concerns and NDAs, I am unable to disclose detailed information about the data and its features, limiting the depth of the analysis. Future research should focus on analysing and enhancing model interpretability, which is crucial for understanding model decisions in sensitive areas such as transaction monitoring. Additionally, research should examine the application of these models to later stages of transaction monitoring, like the investigation phase. This approach would extend our understanding of the models' practical utility and also increase their applicability in enhancing money laundering detection.

Overall, this study further evaluates the deep learning models using a real dataset and extends their application to a new task within transaction monitoring. I identify optimal use cases for specific models tailored to tasks and datasets, enhancing targeted strategies in transaction monitoring. The findings highlight the advantages that transformer-based models bring to this domain, encouraging continued research into these models to further their development and application.

Chapter 7

Conclusion and Future Work

7.1 Summary

This research aimed to investigate advanced machine learning and deep learning techniques to enhance the detection of money laundering activities in financial institutions, specifically within transaction monitoring. To address the limitations of traditional rule-based systems, I pursued several key objectives: (i) to comprehensively review and critically evaluate the current state of transaction monitoring technologies, (ii) to analyse the insights and expectations of anti-money laundering (AML) specialists regarding industry challenges and the potential of advanced machine learning approaches, (iii) to develop and assess a synthetic AML transaction data generator to overcome data scarcity and sensitivity issues, (iv) to develop and evaluate a transformer-based approach for transaction monitoring detection, and to evaluate the developed model further using a real dataset (v) at a later phase in the transaction monitoring process. This conclusion summarises how each of these objectives has been realised and discusses the implications of our findings for the field of AML.

Chapters 2 and 3 addressed objectives (i) and (ii) through a comprehensive systematic literature review that identified critical gaps and shortcomings in existing research, facilitating targeted advancements. This was complemented by semi-structured interviews with AML specialists, who provided insights into the field's current challenges, requirements, and potential solutions. The systematic literature review highlighted several key gaps that guided this study: insufficient academic-industry collaboration, limited data availability, inadequate research on deep learning, and a lack of experiments using real data. Overall, Chapter 2 findings aid stakeholders in selecting strategies to address challenges and direct future research in transaction monitoring, enhancing knowledge and decision-making. Chapter 3 focused on bridging academia and the practical needs of the industry. Key findings emphasised the need to explore machine learning and deep learning techniques that analyse transactions at both group and individual levels to enhance detection capabilities across various patterns. The study also advocated for the integration of feedback loops into transaction monitoring systems and underscored the importance of high-quality datasets. Additionally, I identified strategies to enhance traditional rule-based methods with scorecard techniques, reducing false positives. These contributions provide valuable tools and insights for future research, helping stakeholders combat financial crime more effectively.

Chapter 4 addresses the challenge of data scarcity by developing a synthetic dataset specifically to advance research in transaction monitoring (Objective iii). The dataset creation process involved broadening the spectrum of normal and suspicious transaction typologies, and incorporating features such as geographic locations and connections to high-risk countries, adding complexity. Additional typologies such as 'Structuring', 'Smurfing', and 'Deposit-Send', among others were created based on results from the semi-structured interviews. This enriched dataset was then evaluated using traditional machine learning algorithms. To assess its effectiveness, the created dataset was compared to the most utilised synthetic dataset, AMLSim. The best performing model, Random Forest, attained a ROC-AUC score of 89.83% on SAML-D, whereas it achieved a higher score of 96.39% on AMLSim. All evaluated models demonstrated superior performance on the AMLSim datasets compared to the SAML-D dataset, showcasing the SAML-D datasets' higher complexity. The experimental results highlighted a reduced effectiveness in detecting money laundering typologies compared to the existing dataset, demonstrating the increased challenge the SAML-D dataset presents. These findings underscore its significant potential as a tool for developing and testing more sophisticated detection techniques in the field.

Finally, Chapters 5 and 6 of the dissertation explored the development and experimental testing of deep learning models for transaction monitoring, which have not previously been evaluated within this domain. The application of transformer-based models was explored, and a method was developed that outperforms existing models (Objective iv). Experiments were conducted using both the developed synthetic dataset, SAML-D, and a real dataset provided by Bank X (Objective v) to provide a comprehensive evaluation of these models. On the SAML-D dataset, the transformer-based models demonstrated superior performance, with TabAML achieving an ROC-AUC score of 93.01% and TabTransformer achieving 85.94%, both surpassing XGBoost's score of 81.12%, which is known for its exceptional ability in classification tasks within tabular data and in the AML literature. Conversely, on the dataset provided by Bank X, XGBoost emerged as the most effective model, achieving an ROC-AUC score of 88.73%, while TabAML and TabTransformer attained scores of 84.89% and 85.10% respectively. These experiments revealed distinct preferences for model application based on the specific type of transaction data and task: transformer models were more effective at handling interlinked transactions, whereas the XGBoost model excelled in scenarios involving case-based, aggregated features. The utility of the residual attention and shared embedding mechanisms showed that they significantly enhance performance on interlinked transaction data by focusing on intricate dependencies between transactions. However, these benefits did not extend to case-based aggregated features, where the residual attention and shared embedding mechanisms increased processing time without improving outcomes. The key findings underscore the potential of transformer-based models to enhance transaction monitoring for financial institutions and pave the way for future studies to explore and refine the use of advanced deep learning techniques in combating financial crimes.

Additionally, Chapter 6 includes a detailed examination of the feature set using the performing model, XGBoost. This research was motivated by the desire to identify the most crucial features that enable the model to detect money laundering within the dataset, aiming to enhance explainability and understanding. Additionally, it was to determine whether better performance could be achieved by using a subset of features, reducing the computational costs. The study revealed that using the top 40 SHAP-selected features increased the ROC-AUC score by 0.12%, demonstrating that targeted selection of features can boost both the model's performance and computational efficiency, essential for effective transaction monitoring. Also, the significance of feedback data was highlighted in the performance of the XGBoost model when using the 'Risk Indicator' features in isolation.

7.2 Conclusion

This research has successfully addressed its aim of exploring and advancing the application of machine learning and deep learning techniques for transaction monitoring in AML efforts. The conclusions are structured around the research objectives and provide a critical appraisal of the outcomes of this study.

Objective (i): To comprehensively review and critically evaluate the current state of transaction monitoring technologies. This objective was achieved through a systematic literature review, which identified critical gaps in the existing body of research. Key findings include the limited integration of advanced machine learning models, insufficient academic-industry collaboration, and inadequate amount and therefore utilisation of real-world datasets. These insights provided a strong foundation for subsequent research and highlighted the need for more robust and adaptable AML technologies.

Objective (ii): To analyse the insights and expectations of AML specialists regarding industry challenges and the potential of advanced machine learning approaches. This objective was realised through semi-structured interviews with AML professionals. The interviews provided valuable perspectives on the practical challenges faced in the industry, such as high false-positive rates and the limitations of traditional rule-based systems. The findings underscored the industry's openness to adopting advanced analytics, particularly models capable of learning from dynamic transaction behaviours and incorporating feedback mechanisms.

Objective (iii): To develop and assess a synthetic AML transaction data generator, aimed at overcoming the challenges related to the scarcity and sensitivity of transaction data for research. The synthetic dataset SAML-D, was developed to introduce greater complexity compared to existing datasets such as AMLSim. SAML-D effectively challenged detection models, revealing the need for more sophisticated techniques to address complex money laundering typologies. The evaluated detection models achieved lower performance metrics when evaluated on SAML-D compared to the existing datasets. This outcome reflects the increased complexity and difficulty presented by SAML-D, underscoring its potential as a robust tool for advancing AML research. Although this study makes significant contributions towards this objective, further research could focus on leveraging real transaction data to develop even more complex and realistic synthetic datasets, enhancing the applicability and generalisability of AML detection models.

Objective (iv): To develop and evaluate a transformer-based approach for transaction monitoring detection. A transformer-based model, Tab-AML, was developed and evaluated, with its performance compared against the existing TabTransformer and TabNet models. Tab-AML demonstrated superior performance on the synthetic SAML-D dataset, outperforming both transformer-based and traditional algorithms such as XGBoost, particularly in scenarios involving interlinked transactions. This highlighted its enhanced capability to capture complex relational patterns. Additionally, the integration of residual attention and shared embedding mechanisms significantly improved Tab-AML's performance, underscoring the value of deep learning techniques in advancing transaction monitoring within AML applications.

Objective (v): To evaluate the developed model further using a real dataset at a later phase in the transaction monitoring process. The transformer-based and traditional models were evaluated using real transaction data from Bank X, providing critical insights into their practical applicability. Interestingly, XGBoost outperformed the deep learning models in scenarios involving aggregated case-based features. This finding highlights the importance of model selection based on the specific characteristics of the dataset and the transaction monitoring phase. The analysis also revealed that while transformer models excel in identifying intricate dependent transaction patterns, traditional models remain effective for aggregated data.

7.3 Limitations and Future Work

Building upon this research, several directions can further enhance AML transaction monitoring by addressing the identified limitations and leveraging advanced machine learning and deep learning techniques.

One limitation is the scope of the collaboration, which primarily involved specialists from the banking sector within the UK. The limited diversity of participants, particularly from regulatory bodies and emerging sectors like cryptocurrency, may constrain the generalisability of the findings across different financial environments. To address this limitation, future studies could involve a more diverse group of AML specialists. This broader collaboration can provide a more comprehensive understanding of current challenges and requirements, ensuring models are robust and capable of addressing a wide spectrum of real-world scenarios.

Another limitation lies in the synthetic data generation process. Although the development of SAML-D enhanced existing synthetic datasets, the diversity and complexity of simulated transaction scenarios remain limited compared to real-world conditions. Future work could build upon SAML-D by incorporating more diverse and complex transaction scenarios, such as transactions involving shell companies, politically exposed persons, and trade-based money laundering. Future work could also explore techniques such as generative adversarial networks (GANs) to generate data by learning patterns from real transactions. GANs achieve this through adversarial training, where a 'Generator' creates realistic fake transactions and a 'Discriminator' distinguishes them from real ones, continuously improving the quality of the synthetic data. This approach could produce high quality data that mimics complex laundering behaviours, enhancing model development and data augmentation while preserving data privacy.

Despite achieving strong results with the transformer-based models, challenges remain in model optimisation. Hyperparameter tuning was done through a manual systematic approach, tailored to address computational constraints. While this method produced meaningful results, it restricted the ability to explore a wider range of hyperparameter configurations that could potentially enhance the model's performance. Future research could explore comprehensive hyperparameter optimisation strategies, such as using automated tools like Optuna, to maximise model performance.

The lack of exploration into the explainability and interoperability of the model's outcomes in this study is another limitation that warrants further research. The complexity of attention mechanisms can obscure transparency, which is crucial for regulatory compliance and fostering investigator trust in AML systems. Future research should prioritise the development of techniques that enhance the explainability of attention mechanisms within transformer architectures, ensuring that decision-making processes become more transparent and interpretable. Integrating advanced technologies, such as large language models (LLMs) and AI agents, with tools like Tab-AML could significantly improve the investigation process. These technologies can assist in analysing and contextualising model outputs, presenting insights in a more intuitive and accessible manner. By providing investigators with clearer justifications for flagged transactions, this integration can enhance the efficiency of alert reviews and support well-informed decisions on whether to file a SAR.

Future research can address these limitations and contribute to more robust tools for combating financial crime. Advancements in model optimisation, data generation, interdisciplinary collaboration, and practical implementation will collectively improve the integrity of financial institutions, aligning with the broader goal of enhancing global financial security.

References

Akiba, Takuya et al. (2019). “Optuna: A Next-Generation Hyperparameter Optimization Framework”. In: *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2623–2631.

Alarab, Ismail, Simant Prakoonwit, and Mohamed Ikbal Nacer (2020). “Competence of Graph Convolutional Networks for Anti-Money Laundering in Bitcoin Blockchain”. In: *Proceedings of the 2020 5th International Conference on Machine Learning Technologies*. ICMLT ’20. Beijing, China: Association for Computing Machinery, pp. 23–27. ISBN: 9781450377645. DOI: 10.1145/3409073.3409080.

Alexandre, Claudio Reginaldo and João Balsa (2023). “Incorporating machine learning and a risk-based strategy in an anti-money laundering multiagent system”. In: *Expert Systems with Applications* 217, p. 119500. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2023.119500.

Ali, Ahmed Hussein (2019). “A Survey on Vertical and Horizontal Scaling Platforms for Big Data Analytics”. In: *International Journal of Integrated Engineering* 11.6, pp. 138–150.

Alshantti, Abdallah and Adil Rasheed (2021). “Self-Organising Map Based Framework for Investigating Accounts Suspected of Money Laundering”. In: *Frontiers in Artificial Intelligence* 4, pp. 1–15. ISSN: 2624-8212. DOI: 10.3389/frai.2021.761925.

Alsuwailem, Abdullah A.S. and Akhtar Jamal Khan J. Saudagar (2020). “Anti-money laundering systems: a systematic literature review”. In: *Journal of Money Laundering Control* 23.4, pp. 833–848. DOI: 10.1108/JMLC-02-2020-0018.

Altman, E. et al. (2023). *Realistic Synthetic Financial Transactions for Anti-Money Laundering Models*. arXiv preprint. Available from: https://arxiv.org/abs/2306.16424 [Accessed 7 Mar. 2024].

Altmann, André et al. (2010). “Permutation importance: a corrected feature importance measure”. In: *Bioinformatics* 26.10, pp. 1340–1347. DOI: 10.1093/bioinformatics/btq134.

Arik, Sercan Ö. and Tomas Pfister (2021). “TabNet: Attentive Interpretable Tabular Learning”. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 35.8, pp. 6679–6687. DOI: 10.1609/aaai.v35i8.16826.

Arora, C., M. Sabetzadeh, and L.C. Briand (2019). “An empirical study on the potential usefulness of domain models for completeness checking of requirements”. In: *Empirical Software Engineering* 24, pp. 2509–2539. DOI: 10.1007/s10664-019-09693-x.

Astrova, Irina (2023). “Anti-money Laundering Powered by Graph Machine Learning: ”Show Me Your Friends and I Will Tell You Who You Are””. In: *Intelligent Decision Technologies*, pp. 1–19. DOI: 10.3233/IDT-220193.

Ba, Jimmy, Jamie Ryan Kiros, and Geoffrey E. Hinton (2016). *Layer Normalization*. arXiv preprint. Available from: https://arxiv.org/abs/1607.06450 [Accessed 17 Mar. 2024].

Baldi, Pierre (July 2012). “Autoencoders, Unsupervised Learning, and Deep Architectures”. In: *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*. Vol. 27. Proceedings of Machine Learning Research. PMLR, pp. 37–49.

El-Banna, M. M., M. H. Khafagy, and H. M. El Kadi (2020). “Smurf Detector: a Detection technique of criminal entities involved in Money Laundering”. In: *2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE)*. Aswan, Egypt, pp. 64–71.

Betron, M. (2012). “The state of anti-fraud and AML measures in the banking industry”. In: *Computer Fraud & Security* 2012.5, pp. 5–7.

Bishop, Christopher M. and Nasser M. Nasrabadi (2006). *Pattern Recognition and Machine Learning*. Vol. 4. New York: Springer, p. 738.

Bisong, Ekaba (2019). “Logistic Regression”. In: *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Berkeley, CA: Apress. DOI: 10.1007/978-1-4842-4470-8_20.

Bolton, Richard J. and David J. Hand (2002). “Statistical Fraud Detection: A Review”. In: *Statistical Science* 17.3, pp. 235–255. DOI: 10.1214/ss/1042727940.

Borisov, Vadim et al. (2022). “Deep Neural Networks and Tabular Data: A Survey”. In: *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–21. DOI: 10.1109/TNNLS.2022.3229161.

Bottou, Léon (2012). “Stochastic Gradient Descent Tricks”. In: *Neural Networks: Tricks of the Trade*. Ed. by Grégoire Montavon, Genevieve B. Orr, and Klaus-Robert Müller. Vol. 7700. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-642-35289-8_25.

Breiman, Leo (2001). “Random Forests”. In: *Machine Learning* 45, pp. 5–32. DOI: 10.1023/A:1010933404324.

Breunig, Markus M. et al. (2000). “LOF: Identifying Density-Based Local Outliers”. In: *SIGMOD Rec.* 29.2, pp. 93–104. DOI: 10.1145/335191.335388.

Butgereit, L. (2021). “Anti Money Laundering: Rule-Based Methods to Identify Funnel Accounts”. In: *2021 Conference on Information Communications Technology and Society (ICTAS)*, pp. 21–26. DOI: 10.1109/ICTAS50802.2021.9394990.

Cai, Jie et al. (2018). “Feature selection in machine learning: A new perspective”. In: *Neurocomputing* 300, pp. 70–79.

Camino, R. D. et al. (2017). “Finding Suspicious Activities in Financial Transactions and Distributed Ledgers”. In: *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 787–796. ISBN: 2375-9259. DOI: 10.1109/ICDMW.2017.109.

Canbek, Gürol et al. (2017). “Binary classification performance measures/metrics: A comprehensive visualized roadmap to gain new insights”. In: *2017 International Conference on Computer Science and Engineering (UBMK)*. IEEE, pp. 821–826.

Canhoto, Ana Isabel (2021). “Leveraging machine learning in the global fight against money laundering and terrorism financing: An affordances perspective”. In: *Journal of Business Research* 131, pp. 441–452. ISSN: 0148-2963. DOI: 10.1016/j.jbusres.2020.10.012.

Chandrashekar, Girish and Ferat Sahin (2014). “A survey on feature selection methods”. In: *Computers & Electrical Engineering* 40.1, pp. 16–28. ISSN: 0045-7906. DOI: 10.1016/j.compeleceng.2013.11.024.

Chang, Chih-Chung and Chih-Jen Lin (2011). “LIBSVM: A library for support vector machines”. In: *ACM Transactions on Intelligent Systems and Technology* 2.3, pp. 1–27. ISSN: 2157-6904. DOI: 10.1145/1961189.1961199.

Chau, Derek and Maarten van Dijck Nemcsik (2020). *Anti-Money Laundering Transaction Monitoring Systems Implementation: Finding Anomalies*. John Wiley & Sons.

Chen, Gang and Jin Chen (2015). “A novel wrapper method for feature selection and its applications”. In: *Neurocomputing* 159, pp. 219–226. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2015.01.070.

Chen, S. H. and R. Venkatachalam (2017). “Agent-based modelling as a foundation for big data”. In: *Journal of Economic Methodology* 24.4, pp. 362–383.

Chen, Tianqi and Carlos Guestrin (2016). “XGBoost: A Scalable Tree Boosting System”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ’16. San Francisco, California, USA: Association for Computing Machinery, pp. 785–794. ISBN: 9781450342322. DOI: 10.1145/2939672.2939785.

Chen, Z., L. Dinh Van Khoa, et al. (2014). “Exploration of the effectiveness of expectation maximization algorithm for suspicious transaction detection in anti-money laundering”. In: *2014 IEEE Conference on Open Systems (ICOS)*, pp. 145–149. DOI: 10.1109/ICOS.2014.7042645.

Chen, Z., L. D. Van Khoa, E. N. Teoh, et al. (2018). “Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review”. In: *Knowledge and Information Systems* 57, pp. 245–285. DOI: 10.1007/s10115-017-1144-z.

Cheng, X. et al. (2021). “Combating emerging financial risks in the big data era: A perspective review”. In: *Fundamental Research* 1.5, pp. 595–606.

Cherrington, M. et al. (2019). “Feature Selection: Filter Methods Performance Challenges”. In: *2019 International Conference on Computer and Information Sciences (ICCIS)*, pp. 1–4. DOI: 10.1109/ICCISci.2019.8716478.

Collin, Matthew, Samantha Cook, and Kimmo Soramaki (2017). *The Impact of Anti-Money Laundering Regulation on Payment Flows: Evidence from SWIFT Data*. Tech. rep. 445. Center for Global Development, p. 52.

Cortes, C. and V. Vapnik (1995). “Support-Vector Networks”. In: *Machine Learning* 20.3, pp. 273–297. DOI: 10.1023/A:1022627411411.

Creswell, John W. and J. David Creswell (2018). *Research Design*. 5th ed. SAGE Publications.

Dahouda, M. K. and I. Joe (2021). “A Deep-Learned Embedding Technique for Categorical Features Encoding”. In: *IEEE Access* 9, pp. 114381–114391. DOI: 10.1109/ACCESS.2021.3104357.

De Koker, L. (2006). “Money laundering control and suppression of financing of terrorism: Some thoughts on the impact of customer due diligence measures on financial exclusion”. In: *Journal of Financial Crime* 13.1, pp. 26–50. DOI: 10.1108/13590790610641206.

Dehouck, Maja and Marieke de Goede (2021). *Public-Private Financial Information-Sharing Partnerships in the Fight Against Terrorism Financing*. University of Amsterdam. 6-30.

Desrousseaux, R., G. Bernard, and J. J. Mariage (2021). “Profiling Money Laundering with Neural Networks: a Case Study on Environmental Crime Detection”. In: *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, pp. 364–369. ISBN: 2375-0197. DOI: 10.1109/ICTAI52525.2021.00059.

Dicicco-Bloom, B. and B. F. Crabtree (2006). “The qualitative research interview”. In: *Medical Education* 40, pp. 314–321. DOI: 10.1111/j.1365-2929.2006.02418.x.

Dill, A. (2021). *Anti-money laundering regulation and compliance: Key Problems and Practice Areas*. Cheltenham, UK: Edward Elgar Publishing Limited.

Domashova, Jenny and Natalia Mikhailina (2021). “Usage of machine learning methods for early detection of money laundering schemes”. In: *Procedia Computer Science* 190, pp. 184–192. ISSN: 1877-0509. DOI: 10.1016/j.procs.2021.06.033.

Eifrem, Emil (2019). “How graph technology can map patterns to mitigate money-laundering risk”. In: *Computer Fraud & Security* 2019.10, pp. 6–8. DOI: 10.1016/S1361-3723(19)30105-8.

Europol (2017). *From Suspicion to Action: Converting Financial Intelligence into Greater Operational Impact*. Europol. Available from: https://www.europol.europa.eu/cms/sites/default/files/documents/ql-01-17-932-en-c_pf_final.pdf [Accessed 9 Mar. 2024].

FATF (2023). *International Standards on Combating Money Laundering and the Financing of Terrorism & Proliferation*. Financial Action Task Force. Available from: http://www.fatf-gafi.org/recommendations.html [Accessed 12 May. 2023].

Fawcett, Tom (2006). “An introduction to ROC analysis”. In: *Pattern Recognition Letters* 27.8, pp. 861–874. ISSN: 0167-8655. DOI: 10.1016/j.patrec.2005.10.010.

Ferreira, L. and C. Almeida (2021). “AML Transaction Monitoring Systems: An Overview”. In: *Journal of Financial Crime* 28.2, pp. 355–372.

Financial Action Task Force (FATF) (2023). *High-risk and Other Monitored Jurisdictions - June 2023*. Available from: https://www.fatf-gafi.org/en/publications/High-risk-and-other-monitored-jurisdictions/Increased-monitoring-june-2023.html [Accessed 12 Mar. 2024].

Gal, Michal and Daniel L Rubinfeld (2019). “Data standardization”. In: *New York University Law Review* 94, pp. 737–770.

Gao, H. et al. (2022). “TSMAE: A Novel Anomaly Detection Approach for Internet of Things Time Series Data Using Memory-Augmented Autoencoder”. In: *IEEE Transactions on Network Science and Engineering*. ISSN: 2327-4697. DOI: 10.1109/TNSE.2022.3163144.

Gao, S. et al. (2006). “Intelligent Anti-Money Laundering System”. In: *2006 IEEE Int. Conf. Serv. Oper. Logist., Informatics*. Shanghai, China, pp. 851–856.

Gao, Z. (2009). “Application of Cluster-Based Local Outlier Factor Algorithm in Anti-Money Laundering”. In: *2009 International Conference on Management and Service Science*, pp. 1–4. DOI: 10.1109/ICMSS.2009.5302396.

Granados, Oscar M. and Alejandro Vargas (2022). “The geometry of suspicious money laundering activities in financial networks”. In: *EPJ Data Science* 11.6. DOI: 10.1140/epjds/s13688-022-00318-w.

Guevara, Jorge, Olmer Garcia-Bedoya, and Oscar Granados (2020). “Machine Learning Methodologies Against Money Laundering in Non-Banking Correspondents”. In: *Applied Informatics*. Springer International Publishing, pp. 72–88. DOI: doi.org/10.1007/978-3-030-61702-8_6.

Halter, Emily Marie et al. (2011). *The puppet masters: how the corrupt use legal structures to hide stolen assets and what to do about it*. Tech. rep. Washington, D.C.: World Bank Group.

Hampo, J. A., E. C. Nwokorie, and J. N. Odii (2023). “A Web-Based KNN Money Laundering Detection System”. In: *European Journal of Theoretical and Applied Sciences* 1.4, pp. 277–288. DOI: 10.59324/ejtas.2023.1(4).27.

Han, J., Y. Huang, S. Liu, et al. (2020). “Artificial intelligence for anti-money laundering: a review and extension”. In: *Digital Finance* 2, pp. 211–239. DOI: 10.1007/s42521-020-00023-1.

Hancock, J.T. and T.M. Khoshgoftaar (2020). “Survey on Categorical Data for Neural Networks”. In: *Journal of Big Data* 7.28. DOI: 10.1186/s40537-020-00305-w.

Hand, D.J. (2009). “Measuring classifier performance: a coherent alternative to the area under the ROC curve”. In: *Machine Learning* 77, pp. 103–123. DOI: 10.1007/s10994-009-5119-5.

Hasan, M. M., J. Popp, and J. Oláh (2020). “Current landscape and influence of big data on finance”. In: *Journal of Big Data* 7, p. 21. DOI: 10.1186/s40537-020-00291-z.

Hayble-Gomes, E. (2023). “The use of predictive modeling to identify relevant features for suspicious activity reporting”. In: *Journal of Money Laundering Control* 26.4, pp. 806–830. DOI: 10.1108/JMLC-02-2022-0034.

He, Ping (2010). “A typological study on money laundering”. In: *Journal of Money Laundering Control* 13.1, pp. 15–32.

He, Ruining et al. (2021). *RealFormer: Transformer Likes Residual Attention*. arXiv preprint. Available from: https://doi.org/10.48550/arXiv.2012.11747 [Accessed 14 Mar. 2024].

Helmy, Tamer H. et al. (2016). “Design of a monitor for detecting money laundering and terrorist financing”. In: *Journal of Theoretical and Applied Information Technology* 85.3, pp. 425–436.

Hendriyetty, N. and B.S. Grewal (2017). “Macroeconomics of money laundering: effects and measurements”. In: *Journal of Financial Crime* 24.1, pp. 65–81. DOI: 10.1108/JFC-01-2016-0004.

Hendrycks, Dan and Kevin Gimpel (2016). *Gaussian Error Linear Units (GELUs)*. arXiv preprint. Available from: http://arxiv.org/abs/1606.08415 [Accessed 28 Mar. 2024].

Hervé, Taud and Mas Jean-François (2018). “Multilayer Perceptron (MLP)”. In: *Geomatic Approaches for Modeling Land Change Scenarios*. Cham: Springer, Chapter 27. DOI: 10.1007/978-3-319-60801-3_27.

Hevner, A. et al. (2004). “Design Science in Information Systems Research”. In: *Management Information Systems Research Center*, pp. 75–105.

Hoffer, Elad, Itay Hubara, and Daniel Soudry (2017). “Train longer, generalize better: closing the generalization gap in large batch training of neural networks”. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS’17. Long Beach, California, USA: Curran Associates Inc., pp. 1729–1739. ISBN: 9781510860964.

Huang, D. et al. (2018). “CoDetect: Financial Fraud Detection With Anomaly Feature Detection”. In: *IEEE Access* 6, pp. 19161–19174. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2018.2816564.

Huang, Xin et al. (2020). *TabTransformer: Tabular Data Modeling Using Contextual Embeddings*. arXiv preprint. Available from: https://doi.org/10.48550/arXiv.2012.06678 [Accessed 16 July. 2024].

Hussain, Walayat, José M. Merigó, and Muhammad Raheel Raza (2022). “Predictive intelligence using ANFIS-induced OWAWA for complex stock market prediction”. In: *Interna-*

tional Journal of Intelligent Systems* 37.8, pp. 4586–4611. ISSN: 0884-8173. DOI: 10.1002/int.22732.

Irwin, S. M., A. Raymond Choo, and L. Liu (2012). “Modelling of money laundering and terrorism financing typologies”. In: *Journal of Money Laundering Control* 15.3, pp. 316–335.

Jamshidi, M. B. et al. (2019). “A novel multiobjective approach for detecting money laundering with a neuro-fuzzy technique”. In: *IEEE 16th International Conference on Networking, Sensing and Control (ICNSC)*, pp. 454–458. DOI: 10.1109/ICNSC.2019.8743234.

Johannesson, P. and E. Perjons (2014). *A Method Framework for Design Science Research*. An Introduction to Design Science, pp. 75–89.

Johnson, M. C. and S. R. Kessler (2019). “The art and science of semi-structured interviewing: A comprehensive guide for researchers”. In: *Qualitative Research Journal* 21, pp. 131–147. DOI: 10.1177/1468794119825569.

Jolly, Jasper (2020). *Commerzbank fined by UK watchdog FCA for money laundering failings*. The Guardian. Available from: https://www.theguardian.com/business/2020/jun/17/commerzbank-fined-uk-watchdog-fca-money-laundering-failings [Accessed 4 Apr. 2023].

Jullum, Martin et al. (2020). “Detecting money laundering transactions with machine learning”. In: *Journal of Money Laundering Control* 23.1, pp. 173–186. ISSN: 1368-5201. DOI: 10.1108/JMLC-07-2019-0055.

Jun, Tang and Yin Jian (2005). “Developing an intelligent data discriminating system of anti-money laundering based on SVM”. In: *2005 International Conference on Machine Learning and Cybernetics*. Vol. 6, pp. 3453–3457. ISBN: 2160-1348. DOI: 10.1109/ICMLC.2005.1527539.

Kannan, S. and K. Somasundaram (2017). “Autoregressive-based outlier algorithm to detect money laundering activities”. In: *Journal of Money Laundering Control* 20.2, pp. 190–202. ISSN: 1368-5201. DOI: 10.1108/JMLC-07-2016-0031.

Karim, Md. Rezaul et al. (2024). “Scalable Semi-Supervised Graph Learning Techniques for Anti Money Laundering”. In: *IEEE Access* 12, pp. 50012–50029. DOI: 10.1109/ACCESS.2024.3383784.

Keele, Staffs (2007). *Guidelines for performing systematic literature reviews in software engineering*. Report. Technical report, Ver. 2.3 EBSE Technical Report. EBSE.

Ketenci, Utku Gorkem et al. (2021). “A Time-Frequency Based Suspicious Activity Detection for Anti-Money Laundering”. In: *IEEE Access* 9, pp. 59957–59967. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3072114.

Keyan, L. and Y. Tingting (2011). “An Improved Support-Vector Network Model for Anti-Money Laundering”. In: *2011 Fifth International Conference on Management of e-Commerce and e-Government*, pp. 193–196. DOI: 10.1109/ICMeCG.2011.50.

Khac, N. A. Le and M. Kechadi (2010). “Application of Data Mining for Anti-money Laundering Detection: A Case Study”. In: *IEEE International Conference on Data Mining Workshops*, pp. 577–584. ISBN: 2375-9259. DOI: 10.1109/ICDMW.2010.66.

Khritankov, A. (2021). “Hidden Feedback Loops in Machine Learning Systems: A Simulation Model and Preliminary Results”. In: *Software Quality: Future Perspectives on Software*

Engineering Quality. Vol. 404. Lecture Notes in Business Information Processing. Cham: Springer. DOI: 10.1007/978-3-030-65854-0_5.

Koo, Kyungmo, Minyoung Park, and Byungun Yoon (2024). "A Suspicious Financial Transaction Detection Model Using Autoencoder and Risk-Based Approach". In: *IEEE Access* 12, pp. 68926–68939. DOI: 10.1109/ACCESS.2024.3399824.

Kumari, P. and A. Gupta (2020). "An empirical study on the current status of anti-money laundering compliance and the role of artificial intelligence". In: *Journal of Financial Crime* 27.3, pp. 895–914.

Kute, Dattatray Vishnu et al. (2021). "Deep Learning and Explainable Artificial Intelligence Techniques Applied for Detecting Money Laundering–A Critical Review". In: *IEEE Access* 9, pp. 82300–82317. DOI: 10.1109/ACCESS.2021.3086230.

Labib, Niveen M., Mai A. Rizka, and Ahmed E. M. Shokry (2020). "Survey of Machine Learning Approaches of Anti-money Laundering Techniques to Counter Terrorism Finance". In: *Internet of Things—Applications and Future*. Ed. by Aziza Ghalwash et al. Vol. 114. Lecture Notes in Networks and Systems. Singapore: Springer. DOI: 10.1007/978-981-15-3075-3_5.

Larik, A. S. and S. Haider (2011). "Clustering based anomalous transaction reporting". In: *Procedia Computer Science*. Vol. 3, pp. 606–610. DOI: 10.1016/j.procs.2010.12.101.

Le Borgne, Y.A. et al. (2022). *Reproducible Machine Learning for Credit Card Fraud Detection – Practical Handbook*. Université Libre de Bruxelles. Available from: https://github.com/Fraud-Detection-Handbook/fraud-detection-handbook [Accessed 16 Mar. 2024].

LeCun, Yann A. et al. (2012). "Efficient backprop". In: *Neural Networks: Tricks of the Trade*. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 7700 LECTURE NO: Springer Verlag, pp. 9–48. DOI: 10.1007/978-3-642-35289-8_3.

Leite, Gleidson Sobreira, Adriano Bessa Albuquerque, and Plácido Rogerio Pinheiro (2019). "Application of Technological Solutions in the Fight against Money Laundering - A Systematic Literature Review". In: *Applied Sciences* 9.22, pp. 1–29. DOI: 10.3390/app9224800.

Levi, Michael and Peter Reuter (2006). "Money Laundering". In: *Crime and Justice* 34, pp. 289–375. ISSN: 0192-3234. DOI: 10.1086/501508.

Liashchynskyi, Petro B. and Pavlo Liashchynskyi (2019). *Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS*. arXiv preprint. Available from: https://arxiv.org/abs/1912.06059 [Accessed 21 Jan. 2024].

Ling, Charles X, Jin Huang, and Harry Zhang (2003). "AUC: a better measure than accuracy in comparing learning algorithms". In: *Advances in Artificial Intelligence: 16th Conference of the Canadian Society for Computational Studies of Intelligence, AI 2003, Halifax, Canada, June 11–13, 2003, Proceedings 16*. Springer, pp. 329–341.

Little, Roderick JA and Donald B Rubin (2019). *Statistical analysis with missing data*. Vol. 793. John Wiley & Sons.

Liu, F. T., K. M. Ting, and Z. Zhou (2008). "Isolation Forest". In: *Eighth IEEE International Conference on Data Mining*, pp. 413–422. ISBN: 2374-8486. DOI: 10.1109/ICDM.2008.17.

Liu, R. et al. (2011). "Research on anti-money laundering based on core decision tree algorithm". In: *2011 Chinese Control and Decision Conference (CCDC)*. Mianyang, China, pp. 4322–4325. DOI: 10.1109/CCDC.2011.5968986.

Liu, Xuan and Pengzhu Zhang (2008). “Research on Constraints in Anti-Money Laundering (AML) Business Process in China Based on Theory of Constraints”. In: *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*, pp. 213–213. DOI: 10.1109/HICSS.2008.374.

— (2010). “A Scan Statistics Based Suspicious Transactions Detection Model for Anti-money Laundering (AML) in Financial Institutions”. In: *International Conference on Multimedia Communications*, pp. 210–213. ISBN: 978-0-7695-4136-5. DOI: 10.1109/MEDIACOM.2010.37.

Llugsi, R. et al. (2021). “Comparison between Adam, AdaMax and Adam W optimizers to implement a Weather Forecast based on Neural Networks for the Andean city of Quito”. In: *2021 IEEE Fifth Ecuador Technical Chapters Meeting (ETCM)*. Cuenca, Ecuador, pp. 1–6. DOI: 10.1109/ETCM53643.2021.9590681.

Lokanan, Mark E. (2022). “Predicting Money Laundering Using Machine Learning and Artificial Neural Networks Algorithms in Banks”. In: *Journal of Applied Security Research*. DOI: 10.1080/19361610.2022.2114744.

Lopez-Rojas, E. A. and S. Axelsson (2012). “Money laundering detection using synthetic data”. In: *Annual Workshop of the Swedish Artificial Intelligence Society (SAIS)*. Linköping University: Linköping University Electronic Press.

Louppe, Gilles et al. (2013). “Understanding variable importances in forests of randomized trees”. In: *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1*. NIPS’13. Lake Tahoe, Nevada: Curran Associates Inc., pp. 431–439.

Lundberg, Scott M. and Su-In Lee (2017). “A unified approach to interpreting model predictions”. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS’17. Long Beach, California, USA: Curran Associates Inc., pp. 4768–4777. ISBN: 9781510860964.

Luong, Thang, Hieu Pham, and Christopher D. Manning (2015). “Effective Approaches to Attention-based Neural Machine Translation”. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Ed. by Lluís Màrquez, Chris Callison-Burch, and Jian Su. Lisbon, Portugal: Association for Computational Linguistics, pp. 1412–1421. DOI: 10.18653/v1/D15-1166.

Lv, L. T., N. Ji, and J. L. Zhang (2008). “A RBF neural network model for anti-money laundering”. In: *International Conference on Wavelet Analysis and Pattern Recognition*. Vol. 1, pp. 209–215. ISBN: 2158-5709. DOI: 10.1109/ICWAPR.2008.4635778.

Magomedov, S. et al. (2018). “Anomaly detection with machine learning and graph databases in fraud management”. In: *International Journal of Advanced Computer Science and Applications(IJACSA)* 9.11, pp. 33–38. DOI: 10.14569/IJACSA.2018.091104.

Mahootiha, M. (2020). *Money Laundering Data*. Kaggle. Available from: https://www.kaggle.com/datasets/maryam1212/money-laundering-data [Accessed 05 Feb. 2024].

McDowell, John and Gary Novis (2001). “The consequences of money laundering and financial crime”. In: *Economic Perspectives* 6.2, pp. 6–10.

Mekpor, E.S. (2019). “Anti-money laundering and combating the financing of terrorism compliance: Are FATF member states just scratching the surface?” In: *Journal of Money Laundering Control* 22.3, pp. 451–471. DOI: 10.1108/JMLC-09-2018-0057.

Michalak, K. and J. Korczak (2011). “Graph mining approach to suspicious transaction detection”. In: *Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp. 69–75.

Milaj, Jonida and Carolin Kaiser (Apr. 2017). “Retention of data in the new Anti-money Laundering Directive—‘need to know’ versus ‘nice to know’”. In: *International Data Privacy Law* 7.2, pp. 115–125. ISSN: 2044-3994. DOI: 10.1093/idpl/ipx002.

Mucherino, Antonio, Petraq J. Papajorgji, and Panos M. Pardalos (2009). “k-Nearest Neighbor Classification”. In: *Data Mining in Agriculture*. Vol. 34. Springer Optimization and Its Applications. New York, NY: Springer. DOI: 10.1007/978-0-387-88615-2_4.

Naheem, Mohammed Ahmad (2016). “Money laundering: A primer for banking staff”. In: *International Journal of Disclosure and Governance* 13.2, pp. 135–156. ISSN: 1746-6539.

Narkhede, M. V., P. P. Bartakke, and M. S. Sutaone (2022). “A review on weight initialization strategies for neural networks”. In: *Artificial Intelligence Review* 55, pp. 291–322. DOI: 10.1007/s10462-021-10033-z.

Negrini, M. and R. Riccardi (2018). “Rule-Based Transaction Monitoring for Anti-Money Laundering: A Review of the Current State and Future Perspectives”. In: *Journal of Financial Crime* 25, pp. 417–432.

Oztas, B. et al. (2023a). “Enhancing Anti-Money Laundering: Development of a Synthetic Transaction Monitoring Dataset”. In: *2023 IEEE International Conference on e-Business Engineering (ICEBE)*. Sydney, Australia, pp. 47–54. DOI: 10.1109/ICEBE59045.2023.00028.

— (2023b). “Perspectives from Experts on Developing Transaction Monitoring Methods for Anti-Money Laundering”. In: *2023 IEEE International Conference on e-Business Engineering (ICEBE)*. IEEE. Sydney, Australia, pp. 39–46. DOI: 10.1109/ICEBE59045.2023.00024.

Oztas, Berkan et al. (2022). “Enhancing Transaction Monitoring Controls to Detect Money Laundering Using Machine Learning”. In: *2022 IEEE International Conference on e-Business Engineering (ICEBE)*, pp. 26–28. DOI: 10.1109/ICEBE55470.2022.00014.

Palinkas, L. A. et al. (2015). “Purposeful sampling for qualitative data collection and analysis in mixed-method implementation research”. In: *Administration and Policy in Mental Health and Mental Health Services Research* 42, pp. 533–544. DOI: 10.1007/s10488-013-0528-y.

Pargent, Florian et al. (2022). “Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features”. In: *Computational Statistics*, pp. 1–22. ISSN: 1613-9658. DOI: 10.1007/s00180-022-01207-6.

Paula, Ebberth L. et al. (2016). “Deep Learning Anomaly Detection as Support Fraud Investigation in Brazilian Exports and Anti-Money Laundering”. In: *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 954–960. DOI: 10.1109/ICMLA.2016.0172.

Pedregosa, Fabian et al. (2018). *Scikit-learn: Machine Learning in Python*. arXiv: 1201.0490 [cs.LG].

Pellegrina, Lucia Dalla and Donato Masciandaro (2009). “The Risk-Based Approach in the New European Anti-Money Laundering Legislation: A Law and Economics View”. In: *Review of Law & Economics* 5.2, pp. 931–952. DOI: 10.2202/1555-5879.1422.

Phyu, The Hnin and Surapong Uttama (2023). “Improving Classification Performance of Money Laundering Transactions Using Typological Features”. In: *2023 7th International Confer-*

ence on Information Technology (InCIT), pp. 520–525. DOI: 10.1109/InCIT60207.2023.10413155.

Pieth, M. and G. Aiolfi (2003). “The private sector become active: the Wolfsberg process”. In: *Journal of Financial Crime* 10.4, pp. 359–365. DOI: 10.1108/13590790310808899.

Plaksiy, K., A. Nikiforov, and N. Miloslavskaya (2018). “Applying big data technologies to detect cases of money laundering and counter financing of terrorism”. In: *2018 6th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*. IEEE, pp. 70–77.

Pontes, R. et al. (2022). “Anti-money laundering in the United Kingdom: new directions for a more effective regime”. In: *Journal of Money Laundering Control* 25.2, pp. 401–413. DOI: 10.1108/JMLC-04-2021-0041.

Prado, F. F. and L. A. Digiampietri (2020). “A systematic review of automated feature engineering solutions in machine learning problems”. In: *XVI Brazilian Symposium on Information Systems (SBSI’20)*, pp. 1–7. DOI: 10.1145/3411564.3411610.

Quinlan, J. R. (1986). “Induction of decision trees”. In: *Machine Learning* 1, pp. 81–106. DOI: 10.1007/BF00116251.

Raza, Saleha and Sajjad Haider (2011). “Suspicious activity reporting using dynamic bayesian networks”. In: *Procedia Computer Science* 3, pp. 987–991. ISSN: 1877-0509. DOI: 10.1016/j.procs.2010.12.162.

Riccardi, M. and M. Levi (2018). “Cash, Crime and Anti-Money Laundering”. In: *The Palgrave Handbook of Criminal and Terrorism Financing Law*. Ed. by C. King, C. Walker, and J. Gurulé. Cham: Palgrave Macmillan.

Rocha-Salazar, J. D. J., M. J. Segovia-Vargas, and M. D. M. Camacho-Miñano (2021). “Money laundering and terrorism financing detection using neural networks and an abnormality indicator”. In: *Expert Systems with Applications* 169, pp. 1–15. DOI: 10.1016/j.eswa.2020.114470.

Rocha-Salazar, José-de-Jesús, María-Jesús Segovia-Vargas, and María-del-Mar Camacho-Miñano (2021). “Money laundering and terrorism financing detection using neural networks and an abnormality indicator”. In: *Expert Systems with Applications* 169, p. 114470. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2020.114470.

Rouhollahi, Z. et al. (2021). “Towards Proactive Financial Crime and Fraud Detection through Artificial Intelligence and RegTech Technologies”. In: *The 23rd International Conference on Information Integration and Web Intelligence*, pp. 538–546. DOI: 10.1145/3487664.3487740.

Ruchay, Alexey et al. (2023). “The Imbalanced Classification of Fraudulent Bank Transactions Using Machine Learning”. In: *Mathematics* 11.13, p. 2862. DOI: 10.3390/math11132862.

Rui, Xu and D. Wunsch (2005). “Survey of clustering algorithms”. In: *IEEE Transactions on Neural Networks* 16.3, pp. 645–678. ISSN: 1941-0093. DOI: 10.1109/TNN.2005.845141.

Saar-Tsechansky, M. and F. Provost (2007). “Handling missing values when applying classification models”. In: *Journal of Machine Learning Research* 8, pp. 1625–1657.

Savla, R. and J. Levy (2020). “A Critical Review of AML Transaction Monitoring: Challenges and Opportunities”. In: *Journal of Money Laundering Control* 23, pp. 447–463.

Schmidhuber, Jürgen (2015). “Deep learning in neural networks: An overview”. In: *Neural Networks* 61, pp. 85–117. ISSN: 0893-6080. DOI: 10.1016/j.neunet.2014.09.003.

Schönenberg, Regine and Annette von Schönfeld (2013). *Transnational organized crime: Analyses of a global challenge to democracy*. transcript Verlag. DOI: 10.14361/transcript.9783839424957.

Shokry, Amr Ehab Muhammed, Mohammed Abo Rizka, and Nevine Makram Labib (2020). "Counter terrorism finance by detecting money laundering hidden networks using unsupervised machine learning algorithm". In: *International Conference on e-Learning*, pp. 89–97. DOI: 10.33965/ict_csc_wbc_2020_2020081012.

Shwartz-Ziv, Ravid and Amitai Armon (2022). "Tabular data: Deep learning is not all you need". In: *Information Fusion* 81, pp. 84–90. ISSN: 1566-2535. DOI: https://doi.org/10.1016/j.inffus.2021.11.011.

Simonova, A. (2011). "The risk-based approach to anti-money laundering: problems and solutions". In: *Journal of Money Laundering Control* 14.4, pp. 346–358. DOI: 10.1108/13685201111173820.

Simpson, Andrew (2018). "The role of transaction monitoring in ongoing monitoring: AML compliance programmes in Canada". In: *Journal of Financial Compliance* 2.2. Winter 2018-19, pp. 165–175.

Simser, J. (2013). "Money laundering: emerging threats and trends". In: *Journal of Money Laundering Control* 16.1, pp. 41–54.

Srivastava, Nitish et al. (2014). "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15.56, pp. 1929–1958.

Starnini, M. et al. (2021). "Smurf-Based Anti-money Laundering in Time-Evolving Transaction Networks". In: *Machine Learning and Knowledge Discovery in Databases. Applied Data Science Track*. Vol. 12978. Lecture Notes in Computer Science. Springer. DOI: 10.1007/978-3-030-86514-6_11.

Sterling, S. (2015). "Identifying money laundering: Analyzing suspect financial conduct against the speed, cost, and security of legitimate transactions". In: *Journal of Money Laundering Control* 18.3, pp. 266–292. DOI: 10.1108/JMLC-08-2014-0025.

Stevens, E. A. (2020). "Understanding the inductive approach: Benefits and applications in research". In: *Journal of Research Methods and Applications* 16, pp. 1–18. DOI: 10.1080/15546128.2020.1758346.

Stojanović, Branka et al. (2021). "Follow the Trail: Machine Learning for Fraud Detection in Fintech Applications". In: *Sensors* 21.5. ISSN: 1424-8220. DOI: 10.3390/s21051594.

Sudjianto, Agus et al. (2010). "Statistical Methods for Fighting Financial Crimes". In: *Technometrics* 52.1, pp. 5–19. DOI: 10.1198/TECH.2010.07032.

Suresh, R. M. and R. Padmajavalli (2006). "An overview of data preprocessing in data and web usage mining". In: *2006 1st Int. Conf. Digit. Inform. Manage.* IEEE, pp. 193–198.

Suzumura, T. and H. Kanezashi (2021). *Anti-Money Laundering Datasets: InPlusLab Anti-Money Laundering Datasets*. IBM GitHub Repository. Available from: http://github.com/IBM/AMLSim [Accessed 9 Aug. 2024].

Tai, C. and T. Kan (2019). "Identifying Money Laundering Accounts". In: *International Conference on System Science and Engineering (ICSSE)*, pp. 379–382. ISBN: 2325-0925. DOI: 10.1109/ICSSE.2019.8823264.

Tanha, Jafar et al. (2020). "Boosting methods for multi-class imbalanced data classification: an experimental review". In: *Journal of Big Data* 7.70. DOI: 10.1186/s40537-020-00349-y.

Tatulli, Maria Paola et al. (2023). “HAMLET: A Transformer Based Approach for Money Laundering Detection”. In: *Cyber Security, Cryptology, and Machine Learning: 7th International Symposium, CSCML 2023, Be'er Sheva, Israel, June 29–30, 2023, Proceedings*. Be'er Sheva, Israel: Springer-Verlag, pp. 234–250. ISBN: 978-3-031-34670-5. DOI: 10.1007/978-3-031-34671-2_17.

Tertychnyi, P. et al. (2020). “Scalable and Imbalance-Resistant Machine Learning Models for Anti-money Laundering: A Two-Layered Approach”. In: *Enterprise Applications, Markets and Services in the Finance Industry. FinanceCom 2020. Lecture Notes in Business Information Processing*. Vol. 401. Springer. DOI: 10.1007/978-3-030-64466-6_3.

Thomas, D. R. (2006). “A general inductive approach for analyzing qualitative evaluation data”. In: *American Journal of Evaluation* 27, pp. 237–246. DOI: 10.1177/1098214005283748.

Tundis, A., S. Nemalikanti, and M. Mühlhäuser (2021). “Fighting organized crime by automatically detecting money laundering-related financial transactions”. In: *The 16th International Conference on Availability, Reliability and Security*. Vol. 38, pp. 1–10. DOI: 10.1145/3465481.3469196.

UK Government (2020). *National Risk Assessment of Money Laundering and Terrorist Financing 2020*. UK Government. Available from: https://www.gov.uk/government/publications/national-risk-assessment-of-money-laundering-and-terrorist-financing-2020 [Accessed 22 May. 2024].

Unger, B. and E.M. Busuioc (2007). *The Scale and Impacts of Money Laundering*. Edward Elgar, Cheltenham, UK; Northampton, MA. Available from: http://www.loc.gov/catdir/toc/ecip074/2006035573.html [Accessed 29 Mar. 2024].

United Nations Office on Drugs and Crime (2021). *Global Money Laundering*. United Nations Office on Drugs and Crime. Available from: https://www.unodc.org/ [Accessed 6 Mar. 2023].

Usman, A., N. Naveed, and S. Munawar (2023). “Intelligent Anti-Money Laundering Fraud Control Using Graph-Based Machine Learning Model for the Financial Domain”. In: *Journal of Cases on Information Technology (JCIT)* 25.1, pp. 1–20. DOI: 10.4018/JCIT.316665.

Valbuena, D., P. H. Verburg, and A. K. Bregt (2008). “A method to define a typology for agent-based analysis in regional land-use research”. In: *Agriculture, Ecosystems & Environment* 128.1, pp. 27–36.

Vaswani, Ashish et al. (2017). *Attention Is All You Need*. arXiv preprint. Available from: https://doi.org/10.48550/arXiv.1706.03762 [Accessed 21 May. 2024].

Veyder, F. (2003). “Case study: Where is the risk in transaction monitoring?” In: *Journal of Financial Regulation and Compliance* 11.4, pp. 323–328. DOI: 10.1108/13581980310810606.

Viritha, B., V. Mariappan, and V. Venkatachalapathy (2015). “Combating money laundering by the banks in India: compliance and challenges”. In: *Journal of Investment Compliance* 16.4, pp. 78–95. DOI: 10.1108/JOIC-07-2015-0044.

Vorobyev, Ivan and Anna Krivitskaya (2022). “Reducing false positives in bank anti-fraud systems based on rule induction in distributed tree-based models”. In: *Computers & Security* 120, p. 102786. ISSN: 0167-4048. DOI: https://doi.org/10.1016/j.cose.2022.102786.

Wang, X. and G. Dong (2009). “Research on Money Laundering Detection Based on Improved Minimum Spanning Tree Clustering and Its Application”. In: *Second International Sympo-*

sium on Knowledge Acquisition and Modeling. Vol. 2, pp. 62–64. DOI: 10.1109/KAM.2009.221.

Wronka, C. (2022). "“Cyber-laundering”: the change of money laundering in the digital age". In: *Journal of Money Laundering Control* 25.2, pp. 330–344. DOI: 10.1108/JMLC-04-2021-0035.

Xia, Pingfan et al. (2024). "A Novel Heuristic-Based Selective Ensemble Prediction Method for Digital Financial Fraud Risk". In: *IEEE Transactions on Engineering Management* 71, pp. 8002–8018. DOI: 10.1109/TEM.2024.3385298.

Zavoli, Ilaria and Colin King (2021). "The Challenges of Implementing Anti-Money Laundering Regulation: An Empirical Analysis". In: *The Modern Law Review* 84.4, pp. 740–771. DOI: 10.1111/1468-2230.12628.

Zhang, S., X. Wu, and M. Zhu (2010). "Efficient missing data imputation for supervised learning". In: *9th IEEE International Conference on Cognitive Informatics (ICCI'10)*, pp. 672–679. DOI: 10.1109/COGINF.2010.5599826.

Zhang, Y. and P. Trubey (2019). "Machine Learning and Sampling Scheme: An Empirical Study of Money Laundering Detection". In: *Comput Economics* 54, pp. 1043–1063. DOI: 10.1007/s10614-018-9864-z.

Zhiyuan, Chen et al. (2021). "Variational Autoencoders and Wasserstein Generative Adversarial Networks for Improving the Anti-Money Laundering Process". In: *IEEE Access* 9, pp. 83762–83785. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3086359.

Zolkaflil, S., N. Omar, and S. N. F. Syed Mustapha Nazri (2019). "Implementation evaluation: a future direction in money laundering investigation". In: *Journal of Money Laundering Control* 22.2, pp. 318–326. DOI: 10.1108/JMLC-03-2018-0024.

Appendix A: Ethics

In this appendix, the following documents related to the collection of data from human subjects used in this research are included:

- The ethics checklist for the collection of the data.
- The participant information sheet provided to the human subjects.
- The participant agreement form that each human subject signed before participating in the experiment.
- The semi-structured interview questionnaire.

The risk assessment for the part of this research involving human subjects was found to be low.

Bournemouth University

Research Ethics Checklist

About Your Checklist

Ethics ID	40250
Date Created	04/11/2021 14:37:15
Status	Open
Risk	Low

Researcher Details

Name	Berkan Oztas
Faculty	Faculty of Science & Technology
Status	Postgraduate Research (MRes, MPhil, PhD, DProf, EngD, EdD)
Course	Postgraduate Research - FST
Have you received funding to support this research project?	No
Please list any persons or institutions that you will be conducting joint research with, both internal to BU as well as external collaborators.	Bournemouth University

Project Details

Title	Applying new methods to enhance existing Transaction Monitoring (TM) controls to better detect Anti-Money Loundering (AML) risks and reduce inefficiencies.
Start Date of Project	20/09/2021
End Date of Project	20/09/2024
Proposed Start Date of Data Collection	01/02/2022
Supervisor	Deniz Cetinkaya

Summary - no more than 600 words (including detail on background methodology, sample, outcomes, etc.)

Summary

This research aims to enhance Transaction Monitoring systems in Anti-Money Laundering (AML) to reduce false positives while contributing to regulatory requirements. The reason for the research is that rules-based systems are mainly used by financial institutions to detect money laundering activities, which cannot identify hidden and complex money laundering activities effectively and efficiently, resulting in high false-positive alerts. Therefore, increasing human capital cost and processing time for financial institutions. To cooperate with my industrial supervisor (employee at PwC) and gather transactions data would be the most ideal situation, however, gaining access is difficult. Therefore, dummy data can be used. The artefact to be developed will heavily rely on the type of data that is acquired (i.e. labelled transactions may lead to a supervised approach and could be beneficial in gauging how efficient and successful the artefact is). Furthermore, it will be beneficial to have a large data set to train and test on. Also, unbalanced data can be a problem so synthetic data may need to be used. In addition to the acquired data, publicly available data will be used as a different data set to test the artefact. The main objectives are to identify, 1) why the current detection systems created in literature are not adopted by financial institutions and 2) improvements that can be made to the detection phase of transaction monitoring to reduce false positives while keeping true positives high. Next, developing an approach to enhance detection in Transaction Monitoring for AML, will be done and then demonstrated and evaluated. Interview with AML specialists will be done to understand the importance and need for improvements in transaction monitoring, also, to get expert feedback on the problems that I extract in detection solutions within the literature. A focus group, presenting to specialists, will be held to demonstrate and evaluate the artefact. During the evaluation stage, my artefact will be tested using statistical methods (i.e. AUC).

Ethics Overview

The ethical issues involved in the overall research are the author's participation in money laundering activities and encouraging and giving instructions to others to participate in money laundering. These are not the aims of the research and the author will need to be cautious regarding them and manage research misuse. The main purpose of researching money laundering will be to get a better understanding to produce an anti-money laundering system. If transaction data is acquired from a financial institution the ethical issues will include confidentiality and anonymity of that data. However, if this data cannot be acquired synthetic data or publicly available data can be used. Tests on publicly available data will be done to evaluate the artefact regardless of acquiring data from a financial institution. Interviews and a focus group will be conducted during research and the following must be considered. It is important to make sure an interviewee is happy with the location of the interview (private/public area) as well as being aware of his/her safety when commencing the interview. The interviewee's names should be confidential and permission for any recorded contributions should be used in accordance with the participants wants. Assent should be given in written form. The interviewees should have the right to leave without any reason at any given time. The above will also apply when conducting focus groups. All the data gathered will be confidential and should be stored safely following the BU guidelines.

Filter Question: Does your study involve Human Participants?

Participants

Describe the number of participants and specify any inclusion/exclusion criteria to be used

I will be aiming to interview 10-15 specialists in the Anti-Money Laundering (AML) field. To be included the participant has to be a expert in AML. I will be aiming for 6-8 participants for the focus group.

Do your participants include minors (under 16)?	No
Are your participants considered adults who are competent to give consent but considered vulnerable?	No
Is a Disclosure and Barring Service (DBS) check required for the research activity?	No

Recruitment

Please provide details on intended recruitment methods, include copies of any advertisements.

Recruitment will be conducted by contacting specialists through social media (i.e. LinkedIn) and also sending emails to specialists found by searching company employees and LinkedIn. In addition, the Industrial supervisor, Gokhan Aksu is an AML specialist and can be interviewed and get me in contact with other specialists within the company he works at.

Gokhan a Senior Manager in Financial Crimes Analytics at PwC, is the industrial Supervisor. The relationship between the participants and Gokhan Aksu is that they are colleagues.

Do you need a Gatekeeper to access your participants?	No

Data Collection Activity

Will the research involve questionnaire/online survey? If yes, don't forget to attach a copy of the questionnaire/survey or sample of questions.	No
Will the research involve interviews? If Yes, don’t forget to attach a copy of the interview questions or sample of questions	Yes

Please provide details e.g. where will the interviews take place. Will you be conducting the interviews or someone else?

I will be conducting interviews with the AML specialists. I aim to have the interviews face-to-face at the company address, however, due to covid interviews may be done online, via Microsoft Teams. The interview audio will be recorded, and once the transcripts are done they will be deleted. All interviewees will be anonymised in the work and no private information will be used.

Will the research involve a focus group? If yes, don't forget to attach a copy of the focus group questions or sample of questions.	Yes

Please provide details e.g. where will the focus group take place. Will you be leading the focus group or someone else?

The focus group will involve AML specialists where I will be demonstrating the artefact that I creat and getting their opinion on it. The focus group could be held at the company the participants work at. However, an online focus group could be done due to covid.

Will the research involve the collection of audio materials?	Yes
Will your research involve the collection of photographic materials?	No
Will your research involve the collection of video materials/film?	No
Will any audio recordings (or non-anonymised transcript), photographs, video recordings or film be used in any outputs or otherwise made publicly available?	No
Will the study involve discussions of sensitive topics (e.g. sexual activity, drug use, criminal activity)?	No
Will any drugs, placebos or other substances (e.g. food substances, vitamins) be administered to the participants?	No
Will the study involve invasive, intrusive or potential harmful procedures of any kind?	No
Could your research induce psychological stress or anxiety, cause harm or have negative consequences for the participants or researchers (beyond the risks encountered in normal life)?	No
Will your research involve prolonged or repetitive testing?	No

Consent

Describe the process that you will be using to obtain valid consent for participation in the research activities. If consent is not to be obtained explain why.

Participants will be given a participant information sheet and be asked to sign a Participant Agreement Form. An Agreement Form will contain explicit statements of what taking part in the research project involves and what will become of their data collected. The Participants will be involved in getting a better understanding of AML, its importance and the regulatory pressures.

Do your participants include adults who lack/may lack capacity to give consent (at any point in the study)?	No
Will it be necessary for participants to take part in your study without their knowledge and consent?	No

Participant Withdrawal

At what point and how will it be possible for participants to exercise their rights to withdraw from the study?

It will be made clear to the AML specialists that they can withdraw from the interview at whatever time they want with no reason. This will be included in the Participation agreement and information sheets.

If a participant withdraws from the study, what will be done with their data?

The data aquired from the participant will be deleted and not used.

Participant Compensation

Will participants receive financial compensation (or course credits) for their participation?	No
Will financial or other inducements (other than reasonable expenses) be offered to participants?	No

Research Data

Will identifiable personal information be collected, i.e. at an individualised level in a form that identifies or could enable identification of the participant?	No
Will research outputs include any identifiable personal information i.e. data at an individualised level in a form which identifies or could enable identification of the individual?	No

Storage, Access and Disposal of Research Data

Where will your research data be stored and who will have access during and after the study has finished.

The potential data gathered from the industrial supervisor will be stored on OneDrive for business, and the researcher and the supervisors will be the only people with access to it. The interview and focus group data will also be stored on the business OneDrive and deleted once used. The interview and focus group data may potentially be stored on BU's Online Research Data Repository "BORDaR". The transactional data will not be stored as it is a sensitive piece of data.

Once your project completes, will any anonymised research data be stored on BU's Online Research Data Repository "BORDaR"?	Yes

Dissemination Plans

How do you intend to report and disseminate the results of the study?

Peer reviewed journals,Internal Report,Conference presentation

Will you inform participants of the results?	Yes

If Yes or No, please give details of how you will inform participants or justify if not doing so

An email can be sent out to participants that request the results of the research. If requested, I will send a document that include the results.

Final Review

Are there any other ethical considerations relating to your project which have not been covered above?	No

Risk Assessment

Have you undertaken an appropriate Risk Assessment?	No

Filter Question: Does your study involve the use or re-use of data which will be obtained from a source other than directly from a Research Participant?

Additional Details

Please describe the data, its source and how you are permitted to use it	I will use publicly available synthetic transaction data in my research. I will acquire the data from sources like GitHub and past papers written in the same domain. Also, I will attempt to gain transaction data from my industrial supervisor. The data will be fully

	anonomised and confidential.

Attached documents

Participant Agreement Form Focus Group.docx - attached on 26/11/2021 15:44:07

Participant Agreement Form Interview.docx - attached on 26/11/2021 15:44:11

Participant Information Sheet Focus Group.docx - attached on 26/11/2021 15:44:17

Participant Information Sheet Interviews.docx - attached on 26/11/2021 15:44:18

Bournemouth
University

Participant Information Sheet

The title of the research project

Applying new methods to enhance existing Transaction Monitoring (TM) controls to better detect Anti-Money Laundering (AML) risks and reduce inefficiencies.

Invitation to take part

You are being invited to take part in an interview to discuss transaction monitoring within AML. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information.

What is the purpose of the project?

The aim of this PhD research -

Developing an approach to enhance Transaction Monitoring in Anti-Money Laundering (AML) to reduce false positives while contributing to regulatory requirements.

Your expertise will help the researchers understand the importance and problems of transaction monitoring methods. Also, it will aid in developing a transaction monitoring approach. The interviews will help explore specialists opinions. The aims of this interview is listed below;

1. To identify the problems and challenges in AML transaction monitoring within financial institutions.
2. To identify and get an understanding of the requirements a transactional monitoring system needs to be useful/successful.
3. Opinions on potential solutions.

Why have I been chosen?

The reason you have been contacted is because we want to interview people who have knowledge and experience in the Anti-Money Laundering field. The researcher will aim to recruit 8-15 participants.

Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a participant agreement form. We want you to understand what participation involves, before you make a decision on whether to participate.

If you or any family member have an on-going relationship with BU or the research team, e.g. as a member of staff, as student or other service user, your decision on whether to take part (or continue to take part) will not affect this relationship in any way.

Can I change my mind about taking part?

If you decide to take part, you are still free to stop at any time without giving a reason. No questions will be asked if you stop.

If I change my mind, what happens to my information?

After you decide to withdraw from the study, we will not collect any further information from or about you.

As regards information we have already collected before this point, your rights to access, change or move that information are limited. This is because we need to manage your information in specific ways in order for the research to be reliable and accurate. Withdrawal from the study will be possible before the point the transcript is approved, however, beyond this point it not be possible to withdraw as everything will be anonymised. If the collected data is eligible to delete all the data will be discarded of the interviewee. Further explanation about this is in the Personal Information section below.

What would taking part involve?

The interview will take place in a private room agreed between the researcher and the participant, i.e. private room in the University or in an office in the company the participant works at for an individual interview with the researcher. The interview will last approximately 60 minutes and with permission, we will audio record the interview. We also offer the option of an online interview through Microsoft Teams.

Will I be reimbursed for taking part?

Unfortunately we are unable to pay you for your time.

What are the advantages and possible disadvantages or risks of taking part?

Whilst there are no immediate benefits to you participating in the project, it is hoped that this work will improve the current AML detection systems and reduce the amount of money laundering. Furthermore, we hope this research can help institutions reduce costs and comply with AML regulations, avoiding large fines.

Whilst we do not anticipate any risks to you in taking part in this study, if you feel uncomfortable during the interview, the interviewer will pause for a break, after which you can choose to end the interview or carry on.

What type of information will be sought from me and why is the collection of this information relevant for achieving the research project's objectives?

The information we hope to gain will help the researchers understand the importance of AML, its greatest challenges and the requirements a transactional monitoring system needs to be useful/successful. The information gathered will assist in proving research needs to be carried out in the AML field and how much money laundering effects the world. Information acquired will also assist in improving current AML detection systems that fit in with regulatory requirements. All your personal information will be anonymised and not used in research.

Will I be recorded, and how will the recorded media be used?

The interview will be audio recorded. The audio recordings of your activities made during this research will be used to create a transcript. No other use will be made of them without your written permission, and no one outside the project will be allowed access to the original recordings or transcript. Any personal and company information will be removed before publishing. You have the right to review the transcript for accidentally disclosed information "business secrets", for example. The audio recordings will be deleted as soon as the transcript is finalised.

How will my information be managed?

Bournemouth University (BU) is the organisation with overall responsibility for this study and the Data Controller of your personal information, which means that we are responsible for looking after your information and using it appropriately. Research is a task that we perform in the public interest, as part of our core function as a university.

Undertaking this research study involves collecting and/or generating information about you. We manage research data strictly in accordance with:

- Ethical requirements; and
- Current data protection laws. These control use of information about identifiable individuals, but do not apply to anonymous research data: "anonymous" means that we have either removed or not collected any pieces of data or links to other data which identify a specific person as the subject or source of a research result.

BU's Research Participant Privacy Notice sets out more information about how we fulfil our responsibilities as a data controller and about your rights as an individual under the data protection legislation. We ask you to read this Notice so that you can fully understand the basis on which we will process your personal information.

Research data will be used only for the purposes of the study or related uses identified in the Privacy Notice or this Information Sheet. To safeguard your rights in relation to your personal information, we will use the minimum personally-identifiable information possible and control access to that data as described below.

Publication
You will not be able to be identified in any external reports or publications about the research without your specific consent. Otherwise your information will only be included in these materials in an anonymous form, i.e. you will not be identifiable.

Security and access controls
BU will hold the information we collect about you in hard copy in a secure location and on a BU password protected secure network where held electronically.

Personal information which has not been anonymised will be accessed and used only by appropriate, authorised individuals and when this is necessary for the purposes of the research or another purpose identified in the Privacy Notice. This may include giving access to BU staff or others responsible for monitoring and/or audit of the study, who need to ensure that the research is complying with applicable regulations.

Further use of your information
The information collected about you may be used in an anonymous form to support other research projects in the future and access to it in this form will not be restricted. It will not be possible for you to be identified from this data. To enable this use, anonymised data will be added to BU's online Research Data Repository: this is a central location where data is stored, which is accessible to the public.

Keeping your information if you withdraw from the study
If you withdraw from active participation in the study we will keep information which we have already collected from or about you, if this has on-going relevance or value to the study. This may include your personal identifiable information. As explained above, your legal rights to access, change, delete or move this information are limited as we need to manage your information in specific ways in order for the research to be reliable and accurate. However if you have concerns about how this will affect you personally, you can raise these with the research team when you withdraw from the study.

You can find out more about your rights in relation to your data and how to raise queries or complaints in our Privacy Notice.

Retention of research data

Project governance documentation, including copies of signed **participant agreements**: we keep this documentation for a long period after completion of the research, so that we have records of how we conducted the research and who took part. The only personal information in this documentation will be your name and signature, and we will not be able to link this to any anonymised research results.

Research results:

As described above, during the course of the study we will anonymise the information we have collected about you as an individual. This means that we will not hold your personal information in identifiable form after we have completed the research activities.

You can find more specific information about retention periods for personal information in our Privacy Notice.

We keep anonymised research data indefinitely, so that it can be used for other research as described above.

Contact for further information

If you have any questions or would like further information, please contact Berkan Oztas on the following email address: boztas@bournemouth.ac.uk .

In case of complaints

- Any concerns about the study should be directed to Professor Tiantian Zhang, Bournemouth University by email to researchgovernance@bournemouth.ac.uk.

Finally

If you decide to take part, you will be given a copy of the information sheet and a signed participant agreement form to keep.

Thank you for considering taking part in this research project.

Ref & Version: 0774 & V4
Ethics ID number: 40250
Date: 21/10/2022

Bournemouth University

Participant Agreement Form

Full title of project: Applying new methods to enhance existing Transaction Monitoring (TM) controls to better detect Anti-Money Loundering (AML) risks and reduce inefficiencies.

Name, position and contact details of researcher: Berkan Oztas, PhD researcher, boztas@bournemouth.ac.uk

Name, position and contact details of supervisor: Deniz Cetinkaya, Supervisor, dcetinkaya@bournemouth.ac.uk

To be completed prior to data collection activity

Section A: Agreement to participate in the study

You should only agree to participate in the study if you agree with all of the statements in this table and accept that participating will involve the listed activities.

I have read and understood the Participant Information Sheet (0772) and have been given access to the BU Research Participant Privacy Notice which sets out how we collect and use personal information (https://www1.bournemouth.ac.uk/about/governance/access-information/data-protection-privacy).
I have had an opportunity to ask questions.
I have the right to review the transcript from my interview.
I understand that my participation is voluntary. I can stop participating in research activities at any time without giving a reason and I am free to decline to answer any particular question(s).
I understand that taking part in the research will include the following activity/activities as part of the research:
• being audio recorded during the interview
• my words may be quoted in publications, reports, web pages and other research outputs without using my real name.
I understand that, if I withdraw from the study, I will also be able to withdraw my data from further use in the study **except** where my data has been anonymised (as it cannot be identified) or it will be harmful to the project to have my data removed.
I understand that my data may be included in an anonymised form within a dataset to be archived at BU's Online Research Data Repository.
I understand that my data may be used in an anonymised form by the research team to support other research projects in the future, including future publications, reports or presentations.

	Initial box to agree
I consent to take part in the project on the basis set out above (Section A)	

Name of participant (BLOCK CAPITALS)	Date (dd/mm/yyyy)	Signature
BERKAN OZTAS	21/10/2022	BERKAN
Name of researcher (BLOCK CAPITALS)	Date (dd/mm/yyyy)	Signature

Once a Participant has signed, **please sign 1 copy** and take 2 photocopies:

- Original kept in the local investigator’s file
- 1 copy to be kept by the participant (including a copy of PI Sheet)

Semi-Structured Interview Questionnaire

Applying new methods to enhance existing Transaction Monitoring (TM) controls to better detect Anti-Money Laundering (AML) risks and reduce inefficiencies.

Your comments are high valuable in developing a TM approach.

Aims:

1. To identify the **problems and challenges** in AML transaction monitoring within financial institutions.
2. To identify and get an understanding of the **requirements** a transactional monitoring system needs to be useful/successful.
3. Opinions on potential **solutions**

A	About Interviewee
Name	
Location	
Phone number	
Years of experience	
Background	i.e. Industry/Bank or Consultant or Regulator or Tech vendor
Department	
Role	

1 Please list up to five activities that are difficult to detect today for terrorist financing/money laundering using the current transaction monitoring methods.

	Activities	Risk	Why is it difficult today?	How can it be detected in the future?	Comments
1.					
2.					
3.					
4.					
5.					

2 Please list up to five scenarios that generate high false positive rates when detecting for terrorist financing/money laundering using the current transaction monitoring methods.

	Scenario	Risk	Method/rule used	What's the cause for high FP?	Comments
1.					
2.					
3.					
4.					
5.					

3 On a scale of 0-10, how important are the features listed below for a transaction monitoring method in AML? Please circle the scale below by moving the circles on the right of the scales.

	0	1	2	3	4	5	6	7	8	9	10	
1.Interpretability/Explain-ability of the methods decisions												○
2.Flexibility/adaptability to changes in customers behaviors												○
3.Detection speed of the suspicious activity												○
4. Scalability i.e. the number of transactions it can analyse in an single attempt												○
5. Customer experience												○

4 What do you think are the biggest pain points/problems of the current transaction monitoring methods used for AML?

How does it affect financial institutions?

5 What transaction monitoring methods do you currently utilise or planning to utilise to detect money laundering/terrorism financing?

Do you think current AML transaction monitoring methods are able to adapt to changes in customers behaviours, if not why? (changes from traditional typologies to new ones)

6	What are your thoughts on transactional datasets used for transaction monitoring in AML, i.e. volumes, quality, isolated data? What additional data can be used to enhance the detect of terrorism financing/money laundering? (i.e. social media) How can other controls in the AML space impact and supplement transaction monitoring methods to increase the effectiveness of detection? (i.e. controls : KYC, Ongoing Due Diligence, Sanction Screening)

7	What types of transactions do you consider high AML risk (Ranked in order) and your reasonings (characteristics of the transaction)?

8	What features/capabilities are required from a new transaction monitoring method in AML to produce efficient results and be successfully implementable in the industry?

9	What are your opinions on AI/machine learning as a solution to improve transaction monitoring? How can AI/machine learning be used to detect AML risks? What type of machine learning (i.e. supervised/ unsupervised) do you think is the best approach?

10	How does transaction monitoring in AML effect customers experience? How can transaction monitoring be improved to reduce customer dissatisfaction?

11	Do you have any other pain points of transaction monitoring or ideas that could change the direction/industry of transaction monitoring you would like to discuss that we haven’t already?

Appendix B: Exploratory Data Analysis on SAML-D

This appendix offers a comprehensive overview of the exploratory data analysis (EDA) performed on the SAML-D dataset, utilised in Chapters 5 and 6. The EDA aims to uncover patterns, validate assumptions, and gain insights into the dataset's distribution through various visualisations and statistical analyses. By conducting this analysis, we ensure that the dataset aligns with our expectations.

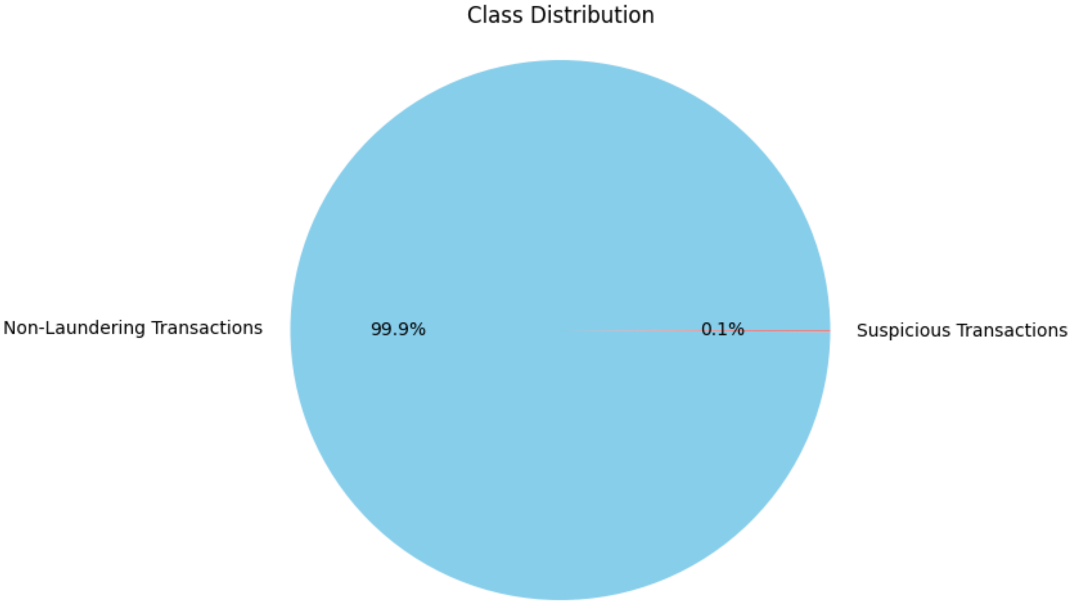


Figure B.1: Class Distribution: This displays the distribution of transactions classified as non-laundering versus suspicious. It illustrates the imbalance in the dataset, with non-laundering transactions making up the majority.

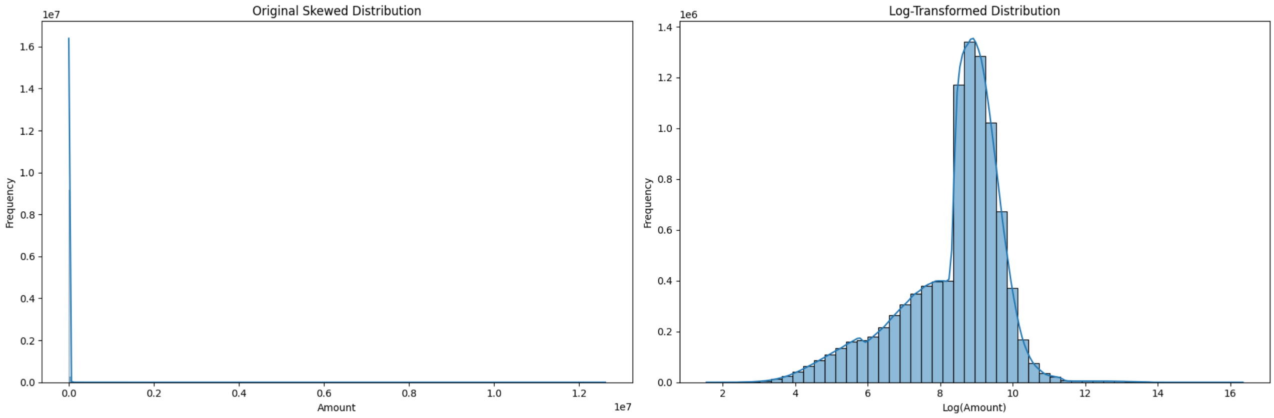


Figure B.2: Distribution of Transaction Amounts: The original and log-transformed distributions of transaction amounts are presented. The log-transformed distribution normalises the data, highlighting the underlying patterns more clearly.

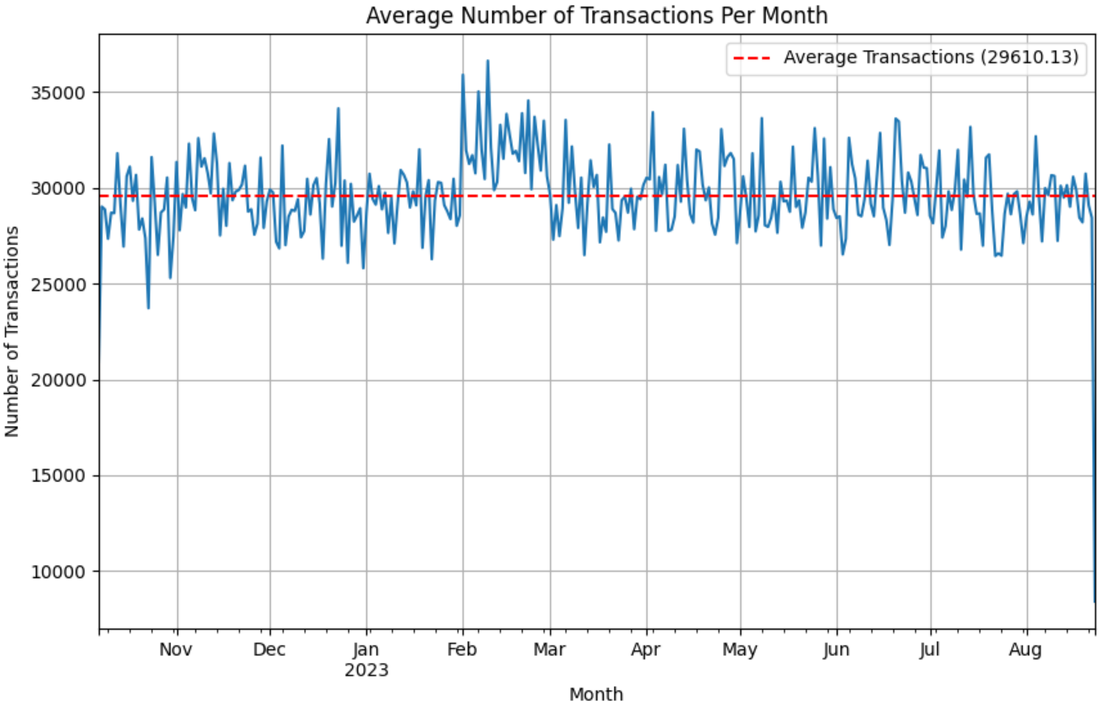


Figure B.3: Average Number of Transactions Per Month: This shows the daily transaction counts over time with an average transaction line. It helps in understanding the trend and seasonality in the transaction data over the examined period.

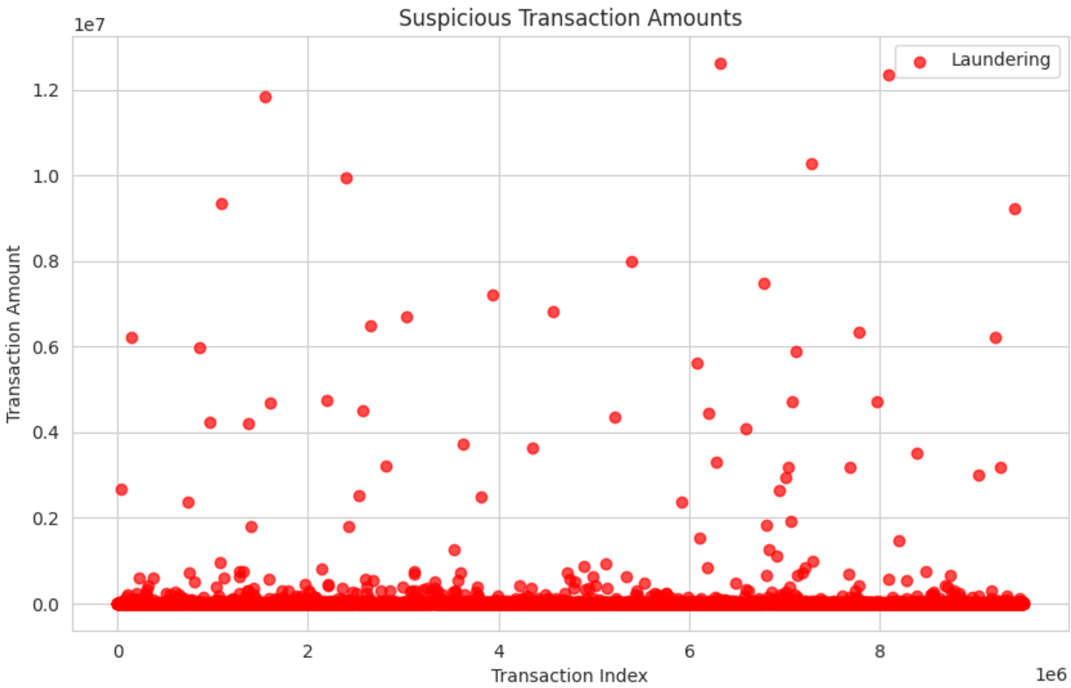


Figure B.4: Suspicious Transaction Amounts: The scatter plot presents the various transaction amounts of suspicious transactions. This visualisation identifies outliers and extreme values in suspicious activities.

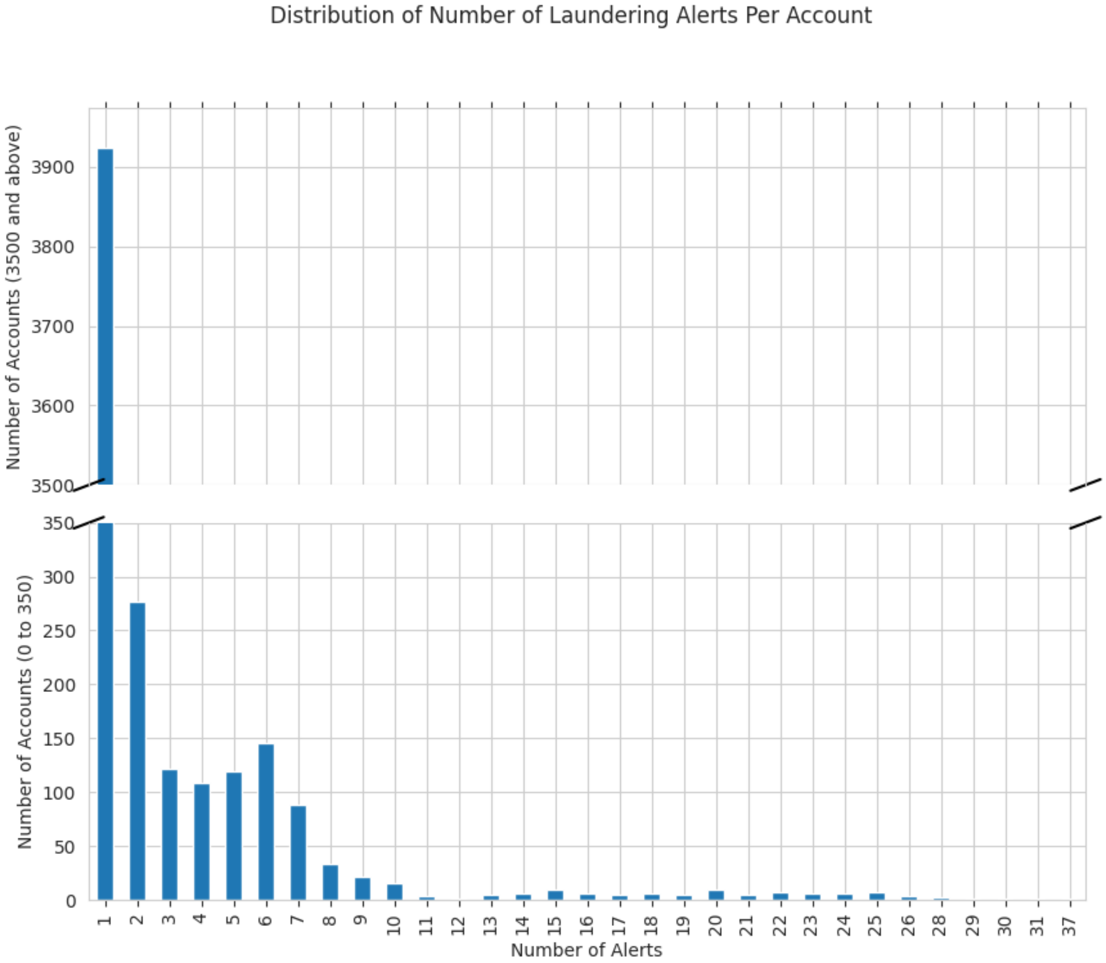


Figure B.5: Distribution of Number of Laundering Alerts Per Account: This histogram reveals the frequency of laundering alerts per account. Most accounts have a low number of alerts, but a small number of accounts have a high number.

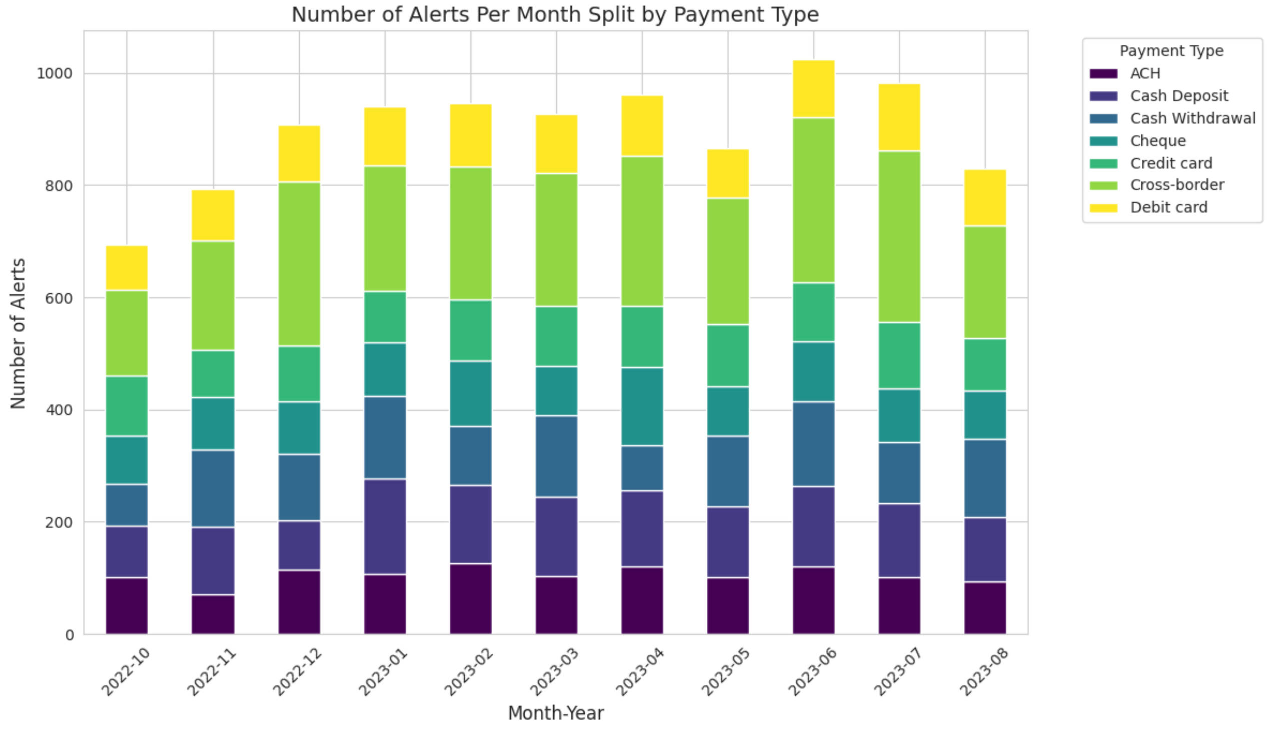


Figure B.6: Number of Alerts Per Month Split by Payment Type: This visualisation aids in understanding which payment methods are most frequently associated with alerts over time.

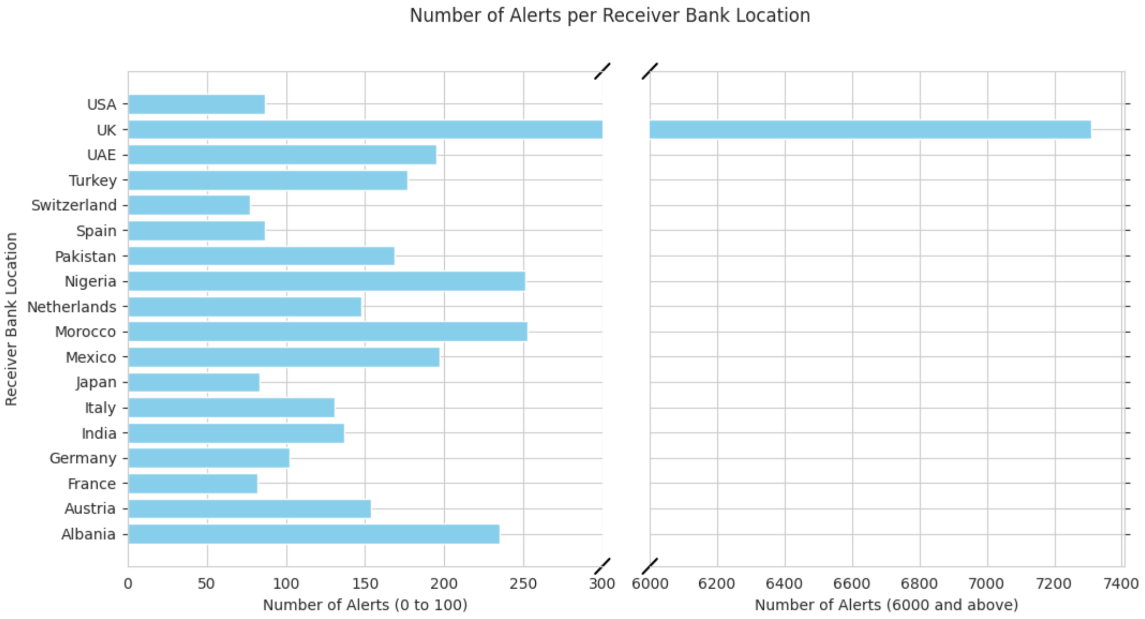


Figure B.7: Number of Alerts per Receiver Bank Location: This information identifies geographical patterns in suspicious activities.

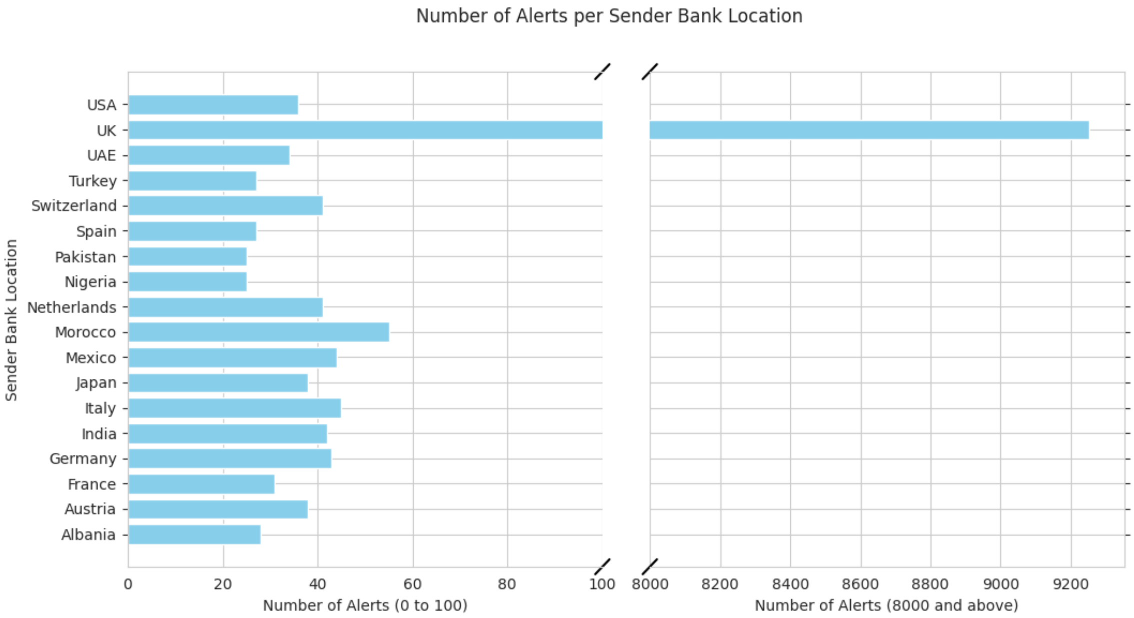


Figure B.8: Number of Alerts per Sender Bank Location: Similar to the receiver bank location, this chart helps in finding the origins of potentially fraudulent transactions.

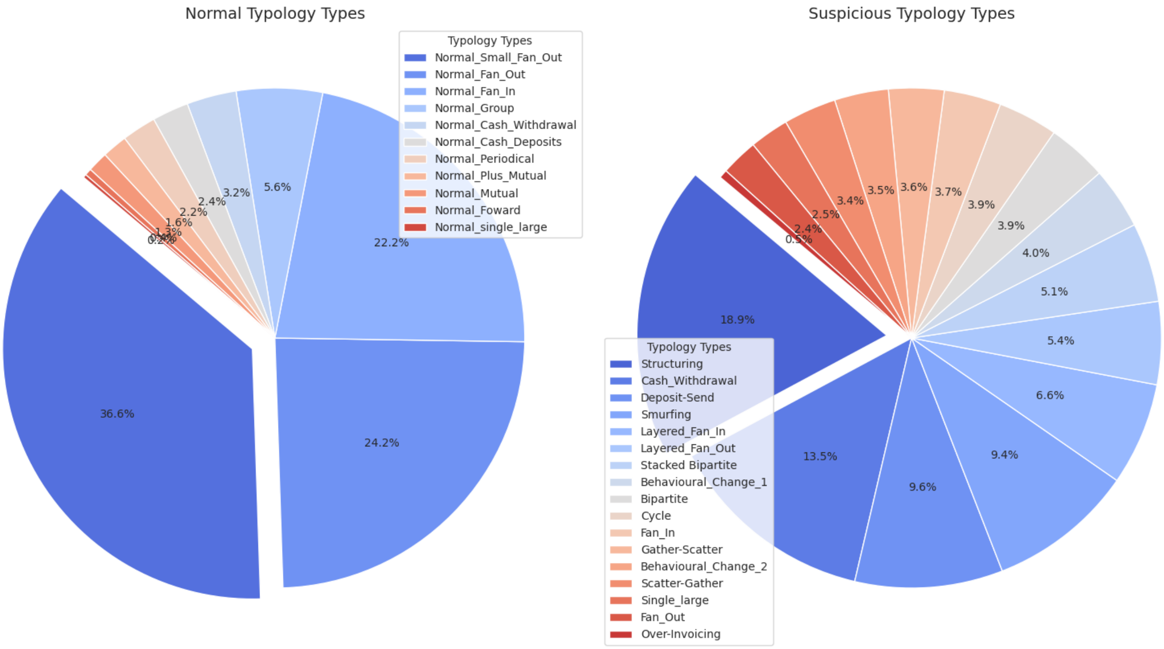


Figure B.9: Typology Types: The pie charts compare the distribution of normal and suspicious transaction typologies. These charts provide insights into the common typologies used in laundering activities versus regular transactions, helping to differentiate between the two.

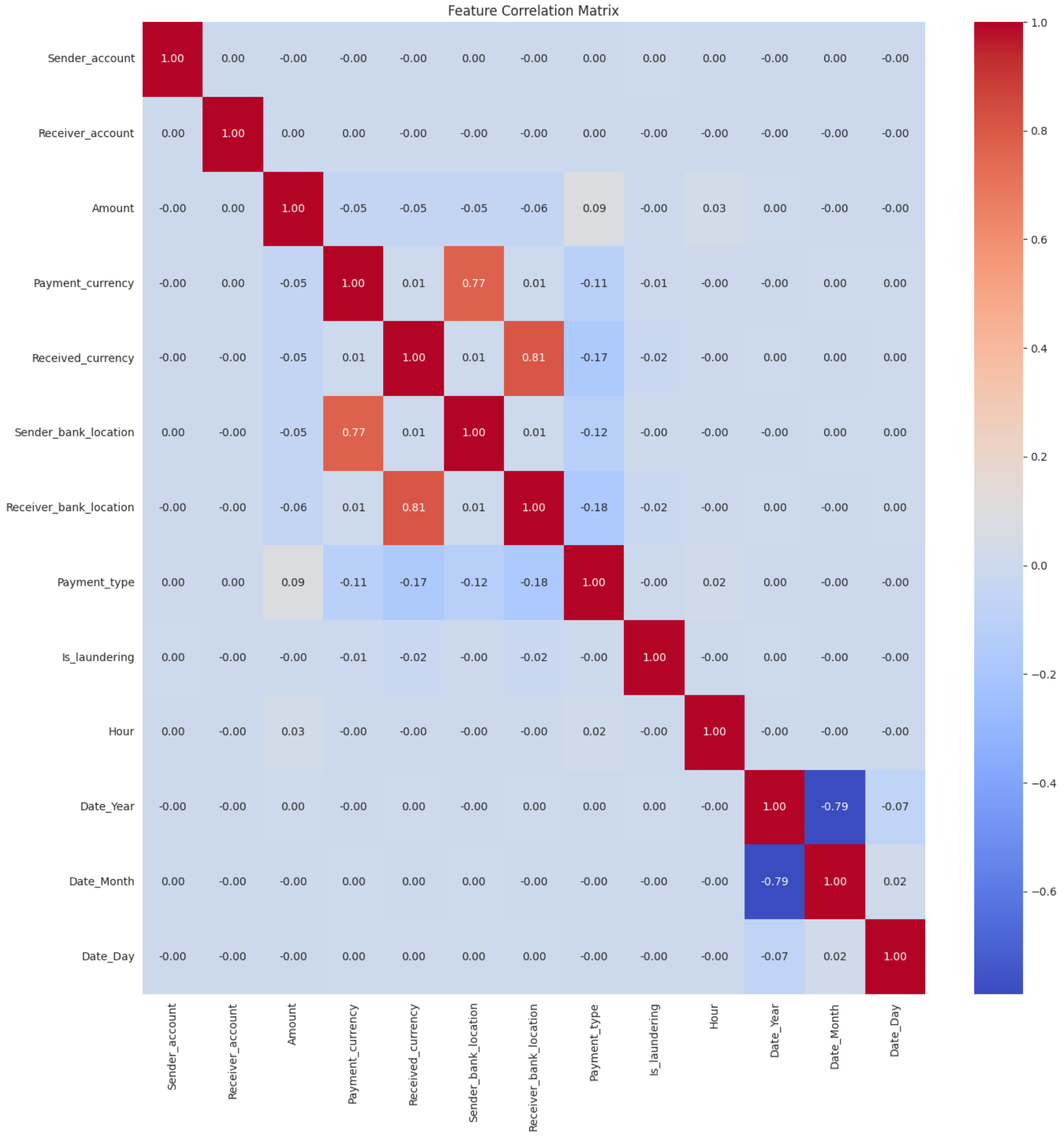


Figure B.10: Feature Correlation Matrix: The heat map of the feature correlation matrix shows the relationships between various features in the dataset. Strong correlations are highlighted.

Each of these visualisations assists in understanding and preparing the data, providing a foundation for the subsequent analysis and model development detailed in the thesis. The insights gained from this EDA confirm that the dataset aligns with expectations and serves as a robust basis for further investigation into transaction patterns and laundering detection.