IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 June 2023, accepted 24 June 2023, date of publication 3 July 2023, date of current version 19 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3291336

RESEARCH ARTICLE

Improving Cancer Detection Classification Performance Using GANs in Breast Cancer Data

EMILIJA STRELCENIA AND SIMANT PRAKOONWIT
Department of Creative Technology, Bournemouth University, BH12 5BB Poole, U.K.
Corresponding author: Emilija Strelcenia (strelceniae@bournemouth.ac.uk)

ABSTRACT Breast cancer is one of the most prevalent cancers in women. In recent years, many studies have been conducted in the breast cancer domain. Previous studies have confirmed that timely and accurate breast cancer detection allows patients to undergo early treatment. Recently, Generative Adversarial Networks have been applied in the medical domain to synthetically generate image and non-image data for diagnosis. However, the development of an effective classification model in healthcare is difficult owing to the limited datasets. To address this challenge, we propose a novel K-CGAN method trained in different settings to generate synthetic data. This study applied five classification methods and feature selection to non-image Wisconsin Breast Cancer data of 357 malignant and 212 benign cases for evaluation. Moreover, we used recall, precision, accuracy, and F1 Score on the synthetic data generated by the K-CGAN model to verify the classification performance of our proposed K-CGAN. The empirical study shows that K-CGAN performed well with the highest stability compared to the other GAN variants. Hence, our findings indicate that the synthetic data generated by K-CGAN accurately represent the original data.

INDEX TERMS Data augmentation, diagnosis, breast cancer, GANs.

I. INTRODUCTION

Breast cancer is among the most common cancers found in women worldwide [1], [2]. According to recent studies [3], [4], breast cancer is the second most common form of cancer, after lung cancer. Therefore, it is a leading malignancy in both the developing and developed countries. This form of cancer occurs when strong cells change in size and begin to evolve chaotically. As a result, a mass of cells known as a tumor develops. There is a dire need for early stage preventive measures to save patients' lives. However, it is imperative to mention that there is a high chance of inaccurate diagnoses owing to human errors and a lack of resources. In recent years, technological advancements in deep learning and computer vision have assisted in automating the means of segregating benign instances from malignant ones with cancerous cells [5].

To detect breast cancer at the initial stage, researchers have introduced mammography. According to previous studies [6], [7], mammography reduces mortality by approximately 40%. However, mammography also has several limitations. Among these limitations, false positive (cancer not present) and under-diagnosed prognostic breast cancer are significant concerns [8]. To overcome these challenges, scientists have used multiple techniques to improve the performance of mammography screening. These techniques include noting two views per breast, double reading, analysis of previous mammograms, and yearly interval screening. However, manual detection using traditional techniques results in high economic costs and strain in a limited number of breast imaging radiologist employees [9].

In addition, regular mammograph screening can assist in the early detection of breast cancer. Nevertheless, common issues such as incorrect negatives, low screening rates, and unnecessary biopsies are observed [10].To deal with these shortcomings, deep learning is a promising option to obtain screening accuracy, lower numbers of incorrect negatives, and unnecessary biopsies. Deep learning models can learn hidden features and correlations that cannot be observed with the naked eye [11].

Recently, the massive success of deep learning in computer vision has been owing to the ease of use of large-scale and

The associate editor coordinating the review of this manuscript and approving it for publication was Vlad Diaconita.

labelled training datasets [12], [13]. However, in various medical imaging fields, the availability of such data is difficult and sometimes impossible owing to privacy reasons. Furthermore, the imbalanced class issue naturally occurs in the medical field, as normal images outnumber those with findings. A common method for dealing with overfitting is to synthetically increase the data size via data augmentation [14].

There are various limitations to biomedical datasets, including; their imbalanced nature. The class imbalance issue occurs because of the uneven distribution of instances associated with noncancerous and cancerous cells. Various approaches have been introduced to address this issue, including under-sampling, oversampling, and hybrid sampling methods [15]. In addition, the general approach of data augmentation has been used in previous studies on computer vision to achieve better performance and class distribution of the model [16]. Many recent studies have applied Generative Adversarial Network (GAN) based data augmentation [17], [18], [19], [20], [21]. Previous studies [22], [23], [24] have shown that the data augmentation method using GANs adapts the training distribution and enhances the performance of classifiers on breast cancer datasets.

In recent years, GANs [17] have been used in the medical domain to synthetically generate computed images. In addition, GANs have been applied to non-image data areas, such as transcriptome data for cancer diagnosis, subtyping, and staging. GAN-based data augmentation has seen rapid progress in the synthetic generation of highly realistic samples [18]. Researchers have implemented GANs in medical imaging, such as computed tomography and magnetic resonance imaging [19], [20]. In addition, GANs have been applied to retinal fundi [21], chest radiography [22], histopathology [23] and liver lesions [24] for data augmentation. In addition to these domains, another important domain is where GANs can be highly successful for the augmentation of data in breast cancer detection in mammograms [25].

In our study, we used Generative Adversarial Networks to synthetically generate breast cancer data to enhance the classification of malignant cases. It is imperative to mention that GANs are a generative method with the concept of game theory, where the Generator and Discriminator try to outperform each other. The role of the Generator is to puzzle the Discriminator. Moreover, on the other hand, the Discriminator acts to discriminate the instances it receives from the original dataset and the Generator. We aimed to generate sufficient synthetically malignant cases to balance the original dataset. We also compared other GAN variants with the proposed method.

This study utilized multiple GAN variants for data augmentation to demonstrate how classification methods work with small tabular medical datasets. The study proposed the Novelty Kullback-Leibler Divergence Conditional GAN (K-CGAN) technique for data augmentation and compared it with other state-of-the-art GAN frameworks (LS-GANs, WGANs, NS-GANs, and SDGs-GANs). We use these GAN frameworks to generate artificial training data to avoid the need for large quantities of medical data. The motivation of our study was to address the limited data training issue in the medical field, which makes it difficult to develop a highly efficient classification framework. Moreover, this study aimed to demonstrate the capability of GANs to improve the performance of classification frameworks in the medical field by offering a novel method for data augmentation in small healthcare datasets.

A. GENERATIVE ADVERSARIAL NETWORKS (GANs) IN BREAST CANCER DOMAIN

Strategies such as under-sampling, over-sampling, and feature selection can deal with the adverse effects that occur from imbalanced source data. In recent years, novel data-augmentation strategies such as Generative Adversarial Networks (GANs) have been used to artificially generate additional data [26]. Generally, GANs are employed to image data and comprise two sub-networks: Generator and Discriminator. The role of the Generator is to generate synthetic samples, whereas the Discriminator is designed to discriminate between fake and real samples [34]. In other words, the function of the Generator is to produce samples with features that the Discriminator cannot separate from real samples, thus enriching the original dataset. Compared to other generative approaches, GANs have a higher computational speed and enhanced sample quality. Therefore, GANs are considered superior to other methods [27].

Additionally, GANs show a lower possibility of overfitting classifier risk and are less vulnerable to the impacts of non-pertinent sample features [28]. Data augmentation using generative models is highly effective because only a specific patch of the entire sample needs to be augmented. The applicability of GANs in mammograms has potential for many reasons. For instance, GANs can overcome the unavailability of significant original datasets. Additionally, public datasets comprise only a small proportion of malignant samples in the general population. Another reason for the applicability of GAN is that they're advantageous and may improve cancer detection that could be used in screening programmes. In general, GANs can help cancer screening programs in many ways. For example, generating synthetic images that can be used to train and validate cancer detection algorithms, identifying anomalies in images that may not show obvious signs of cancer, and generating personalized cancer screening images for each patient. These methods have been shown to be effective in detecting breast cancer, prostate cancer, and other types of cancer. Overall, GANs have the potential to significantly improve the accuracy and effectiveness of cancer screening programs [18].

II. RELATED WORK

In this section, we explore recent studies on machine learning-based data augmentation techniques for breast cancer classification.

The summary of some recent studies is shown below:

In [29], an augmentation framework based on a GAN was proposed. Their study applied feature selection and employed five classification algorithms to medical dataset that were augmented using the least-squares GAN (LS-GAN) method. Empirical research has demonstrated that support vector machines (SVM) perform better than other methods by effectively classifying the data. Their LS-GAN-based data augmentation approach offered an effective solution for improving the performance of classification models in the medical field.

The study [30] introduced an Enhanced Generative Adversarial Network (E-GAN) to solve the imbalanced class challenge. The main objective of their study was to classify non-balanced datasets, such as the Wisconsin breast cancer dataset, with great accuracy. This study converted imbalanced data into balanced data in the pre-processing procedure for this purpose. The pre-processing process consists of data cleaning, normalisation, and transformation using the radius synthetic minority oversampling (R-SMOTE) technique. Moreover, a Deep Convolutional GAN was employed to balance the dataset, generating extra samples under the training dataset. The performance of the proposed method was verified using three datasets, and the performance metrics were examined to verify the capability of the proposed framework. The findings demonstrated that the Breast Cancer Wisconsin dataset attained the highest maximum geometric mean of 08.68, 02.93, and 05.41% and a higher Matthews's correlation coefficient (MCC) than the other methods.

The study [31] used five popular GAN methods to synthetically generate data to train binary classifiers and compared the performance of these classifiers in terms of accuracy with scenarios where only real data are focused. According to this study, GAN-generated data are highly applicable for two reasons: the non-availability of medical data and data privacy regulations. The main objective of this study was to examine how GAN-generated data can enhance the classification accuracy, particularly when the dataset is small. For this purpose, this article introduces a framework that considers an extended dataset with real and synthetic data. The findings revealed that the data generated with advanced GAN methods, such as WGAN-GP, offer better binary classification accuracy with larger and smaller data quantities.

The study in [32] argues that imbalanced class challenges can bias the classifier towards the majority class. This scenario causes a problem for deep learning frameworks, which require diverse and copious data to learn patterns. In this study, the authors used GANs to correct the imbalance class issue in tabular datasets. This study surveyed different studies and investigated the experimental methodologies that have attained machine learning efficacy. The authors of this study noticed that GANs successfully rebalanced tabular datasets.

Similarly, research [33] has examined the scope of deep learning models, such as the conditional generative network (CTGAN) and tabular variational autoencoder (TVAE) to synthetically generate tabular data of breast cancer and assist in its diagnosis of breast cancer. Moreover, their research work also introduced an integrated deep learning model, which comprises the generation of breast cancer data that leads to the classification of breast cancer by employing at deep attention-based model (TabNet). The benchmark breast cancer datasets were used to validate the findings. The empirical findings show that the TVAE model performed exceptionally well. In addition, the TabNet framework performed better than the other deep learning classifiers, with accuracy scores of 96.66% and 82.83% in diagnosis and prognosis, respectively.

In their study, [34] emphasized that biomedical data are difficult to acquire and that obtaining a sufficient sample size is also complex and time-consuming. To solve this problem, they proposed Conditional Generative Adversarial Network (CGAN) based feature generation to synthesize sample datasets with class separability. CGAN was trained on 25% of the five datasets, including the Breast Cancer Wisconsin dataset. The findings confirmed that the synthetically generated datasets had better classification than the real-world datasets. The CGAN-based feature space method generated datasets with the desired class distribution and a higher success rate of the classification algorithms.

In their research, [10] implemented various machine learning classifiers to determine the form of breast cancer in potential patients, and the Wisconsin Breast Cancer Dataset was used to understand the usefulness of these models. Furthermore, they used classification methods, including Decision Tree, Naïve Bayes, KNN, Random Forest, SVM and Logistic Regression, to classify benign and malignant breast cancer. Their findings showed that Random Forest and SVM offer the highest accuracy. However, the authors are required to use a larger dataset in future studies.

According to [35], deep learning has demonstrated promising advancements in cancer classification in mammography. However, issues such as imbalanced classes and scarcity of data are the most prominent barriers to further advances. To address this issue, [35] GANs have been used as a data augmentation method for classification networks. They employed a U-Net-based framework with semi supervised learning. The authors synthesized lesions onto normal-appearing mammogram patches and removed the lesions from the patches where they were present. Their study revealed that considering augmented mammogram patches enhanced the overall performance of the model. Furthermore, they demonstrated that the augmented GAN-based regime yielded an AUC of 0.846, which was superior to the baseline.

Another study, [7] attempted to solve the issue of limited labelled data in breast cancer mammographic image classification. To do so, they selected a limited labelled dataset. The authors introduced a method using DC-GAN to synthetically generate images and augment them with real images to enhance the classification accuracy via a deep Convolutional Neural Network (CNN) algorithm. For this purpose, they

used a dataset consisting of 213 normal and 74 cancerous images. The GAN-based training was conducted with batch size of 32 and 4. They synthetically generated 25 images with 32 batch sizes. According to their study, when comparing the accuracy of synthetically generated 25 images in batch sizes of 4 and 32, the batch size of 32 produced better results, with the accuracy of 87%. To validate the model, three experiments were performed. These experiments were CNN-based performance evaluations with and without DC-GAN, a similarity study, and visual turing testing. The findings of these experiments demonstrate that GAN is a suitable option and offers better solutions to limited dataset challenges.

A DiaGRAM learning solution has been proposed [2] for breast cancer screening and diagnosis. This model uses two main frameworks to attain efficient mammogram diagnosis: (a) the model amalgamates a GAN with a deep classifier to learn traits that advantage both, and (b) transfer learning is employed to acclimatize the model trained with one type of data to another. The authors conducted empirical studies using the INbreast and DDSM datasets. The findings confirmed the excellent performance of DiaGRAM in terms of AUC and accuracy measures compared to previous studies. The proposed method, DiaGRAM, showed transfer learning ability similar to the method trained on the DDSM dataset, and the INbreast dataset demonstrated better performance.

In their study, Wu et al. [18] argued that deep learning frameworks have shown excellent results for breast cancer detection in mammograms. Nevertheless, limited data challenges have constrained the effective performance of these models. To address this challenge, [18] GANs have been used to augment mammogram datasets synthetically. They trained a class conditional GAN to perform contextual infilling for this purpose. Subsequently, the authors synthesized lesions from healthy screening mammograms. The authors-demonstrated that GANs can generate high-quality synthetic mammogram patches. Furthermore, their empirical study evaluated an augmented dataset to enhance breast cancer classification performance. Their study demonstrated that the trained classifier trained with augmented GAN-based training data achieved a superior AUROC compared to the frameworks trained with traditionally augmented datasets.

Furthermore, [36] argued that the poor performance of the classifier model is due to an imbalanced class problem caused by the biased classification of the malignant (majority) class. To address this issue, [36] introduced a novel deep learning-based framework to classify breast cancer imbalanced datasets. The study also investigated the impact of the DC-GAN and the effect of batch normalization on the effectiveness of their transfer network framework. The authors used a DC-GAN for data augmentation of the minority class in the initial phase. The rebalanced dataset with the class distribution was then applied as an input to the proposed deep transfer network. The statistical findings of their study confirmed the validity of the proposed architecture, as it achieved a higher score when compared with other state-of-the-art deep networks.

In addition, [37] they evaluated the performance of two machine learning techniques. The first technique is based on CNNs for classifying histological images into malignant and benign classes to detect breast cancer. The other framework they used was based on extracting handcrafted traits encoded by coding models and trained using SVM. The findings of their study confirmed that CNNs performed better than handcrafted trait-based classifiers.

In another study, [29] proposed applying Generative Adversarial Networks (GANs) as data augmentation method in the medical domain. The objective of the study was to improve the performance, stability and precision of classifier by generating synthetic data that acts as real data. For the said purpose, their study utilized feature selection and used five classification techniques to thirteen datasets. Furthermore, the study augmented these datasets using the Least Square GAN. The assessment of generated instances showed that the SVM model performed better than other data augmentation methods. The proposed method using GAN offers a promising way to improve the performance of classification techniques in the medical domain.

A recent study by [38] proposed a novel technique called Sparse CounteRGAN (SCGAN) to generate counterfactual instances to develop causal association between Clinical Information and Molecular (ICM) features and the treatment responses after Neoadjuvant Systematic Therapy (NST). Their generative framework spots the distribution of original cases and thus, makes sure that the generated cases are realistic. In addition, the study introduced a loss function that regularizes the counterfactual to lessen the distance between the counterfactuals and original cases and the distances among the produced counterfactuals to encourage diversity. Moreover, the study evaluated the proposed model on two benchmark datasets and compared their performance on several methods. The findings of their study show that the SCGAN produces realistic counterfactual cases with minor changes in only few features, thus making it an effective method for meaningful understanding about the causal relation between treatment response and ICM features.

A more recent study by [39] introduced a novel hybrid feature selection method. Their model was aimed to develop an effective feature selection technique and to efficiently classify breast lesions. The study combined relied and Binary Harris Hawk Optimization hybrid method for feature selection. Moreover, techniques such as KNN, SVM, Naïve Bayes and Logistic Regression were selected for the classification. The proposed model was tested on three benchmark dataset: MBCD, WDBC and WBCD. The empirical findings suggest that the proposed hybrid model enhances the performance of all classification techniques on the benchmark datasets. In addition, the hybrid model attained superior results than other methods. This experimental study using hybrid model

achieved better results when compared with other recent studies.

Moreover, [40] introduced a multi-round transfer learning and modified Generative Adversarial Network (MTL-MGAN) for cancer detection. The MTL transfers the information between the source domains and target domains to avoid fatigue search of dataset prioritization between numerous datasets. The process allows maximizing the transferability with a multi-round transfer learning procedure and without negative transfer through customization of loss functions. By doing so, the MGAN produces additional training data and creates intermediate domains to bridge the target domains and source domains. The study utilized ten popular datasets for analysis and evaluation of MTL-MGAN. The findings suggest that the MTL-MGAN method has achieved considerable accuracy compared to previous studies. For all the components of MTL-MGAN, ablations were performed to check the performance of the MTL, the prioritization method, the MGAN and the negative transfer avoidance. The ablation studies also provided remarkable findings to confirm the significance of the components of the method in terms of multiple transfer learning, features, instances, customized loss functions and MGAN.

According to [41], GANs have the ability to learn from training data and produce synthetic data with the same traits as the training data. Based on traits of GAN, the study presented GAN's ability as a method of disease prognosis prediction and developed a prognostic model, Preg-GAN, which is based on conditional GAN (CGAN). The objective of this study was to use Preg-GAN to produce the prognosis prediction outcomes on the basis of patient data. The proposed algorithm added the data as conditions to the training phase. Moreover, conditions were utilized as the input to the generator with noises. Furthermore, the study used the gradient penalty approach and Wasserstein distance to avoid mode collapse during training of the Preg-GAN. The study's findings suggest that Preg-GAN attained good scores with the Area under Curve (AUC) of 0.946 and the average accuracy (ACC) of 90.6%. The experimental findings on the breast cancer dataset demonstrate that the proposed method is a consistent prognosis prediction model.

III. METHODOLOGY

In this section, we present the background of GAN-based classification and a comprehensive introduction to GANs [17], WGAN [42], NS-GAN [43], LS GAN [44], SDG GAN [45] and our proposed K-CGAN model. In addition, this section presents the key procedures for our experimental model.

A. GANS

The Generator G is denoted as G: Z $\rightarrow$ X, where Z is the noise space and X is the data space. The Generator attempts to capture real data distribution. Discriminative model D is denoted by D: X $\rightarrow$ [0, 1]. The Discriminator approximates the probability that the samples are generated by the Generator or from the real data distribution. Both the Generator and the Discriminator models compete with each other in a mini-max game. The value function of the GAN is:

$$min_G max_D V(D,G) = E_{x\sim p_r(x)}[logD(x)] + E_{z\sim pz(Z)}[log(1-D(G(z)))] \quad (1)$$

The "pz" in eq. 1 is the distribution of data over noise input "z". Whereas p_r the data distribution over real data "x". In the training phase, we have to make sure that D's actions over real data are precise by maximizing the first term of eq.1. "$E_{x\sim p_r(x)}[logD(x)]$." On the other hand, when $G(z)$ is fake sample, the D is likely to output a probability $D(G(z))$, near to zero by maximizing the second term in eq. 1, $E_{z\sim pz(Z)}[log(1-D(G(z)))]$. Meanwhile, the G is tasked to increase the probability of D for producing high chances for fake cases, hence the G is trained to minimize $E_{z\sim pz(Z)}[log(1-D(G(z)))]$.

In this way, both the D and G play a mini-max game.

B. WASSERSTEIN GAN

In their study, [42] Wasserstein GAN was proposed to minimize the Earth Mover (EM) distance. The authors used EM distance to learn the probabilistic distributions of a real-world dataset. Their study demonstrated that WGANs can solve the training issues present in traditional GANs.

Wasserstein GAN is a substitute of traditional GAN training. In this novel method, the WGAN paper demonstrated that the stability of learning can be improved and problems such as mode collapse can be avoided. Therefore, in this novel method, the authors provided consequential learning curves helpful for hyperparameter searches and debugging. Moreover, the study provided extensive theoretical efforts focusing the deep associations to various distances between distributions.

The simple GAN extensively utilized Jansen-Shannon divergence (commonly known as GAN loss). The WGAN paper introduced Wasserstein metric as it is smoother value space than Jansen-Shannon divergence. The GAN method has several limitations while running with gradients which can cause training instability. For that reason, Wasserstein distance was used in WGAN paper to address these recurring problems. The WGAN loss function is a smart transformational formula on the basis of Kantorovich-Rubenstein duality.

The loss functions for training the Wasserstein GAN is:

$$W(P_r, P_g) = \sup_{\|f\|_L \leq 1} E_{x\sim P_r}[f(x)] - E_{x\sim P_\emptyset}[f(x)] \quad (2)$$

Here, "sup" means supremum, which measures the upper bound, or simple the maximum values and f is 1-Lipschitz function.

Note here that if we replace "$\|f\|_L \leq 1$" for "$\|f\|_L \leq K$" (K is a constant, let's consider it K-Lipschitz), then we can attain $K.W(p_r, p_z)$. Consequently, if we include a parameter function family $\{f_w\}_{w\in W}$ which are K- Lipschitz for a constant

K then the new form of Wasserstein metric can be:

$$W(p_r, p_g) = \max_{w \in W} \mathbb{E}_{x \sim p_r}[f_w(x)] - \mathbb{E}_{z \sim p_z}[f_w(g_\theta(z))] \quad (3)$$

The above process yield us to calculating $W(p_r, p_\theta)$ to a multiplicative constant. The modified function of WGAN, the discriminator is employed to learn "w" to locate a better "f_w". On the other hand, the loss function is arranged to measure the Wasserstein distance between p_r and p_g.

Figure 1 shows the WGAN's Generator Network, which consists of a series of layers, each of which play a crucial role in the process. Figure 2 presents the WGAN's Discriminator Network, which consists of multiple layers.

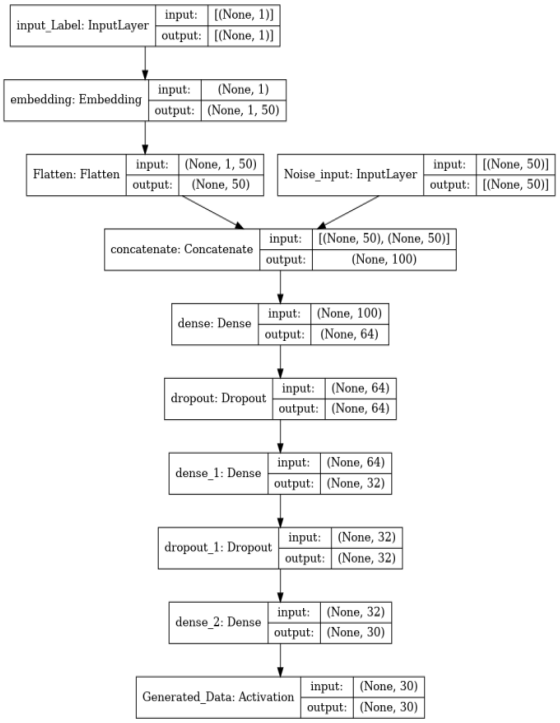


FIGURE 1. The generator network of Wasserstein GAN architecture.

In WGAN, the Discriminator is an opponent network, as it only reports the data distribution that the Generator acquires and does not assess the accuracy of the data. Moreover, weight clipping ensures that the weight of the Generator adheres to the Lipschitz restrictions.

C. NON SATURATING GAN

NS-GAN loss is an alteration of the Generator loss to solve the saturation issue. It is a modification that entails the generator to maximize the log of the Discriminator probabilities in order to generate instances as an alternative of minimizing the log of inverted discriminator probabilities for generated instances.

To get more insight about NS-GAN, we have to understand GAN loss function discussed by [17]. In the standard loss function, the Generator attempt to minimize the function. On the other hand, the Discriminator attempts to maximize

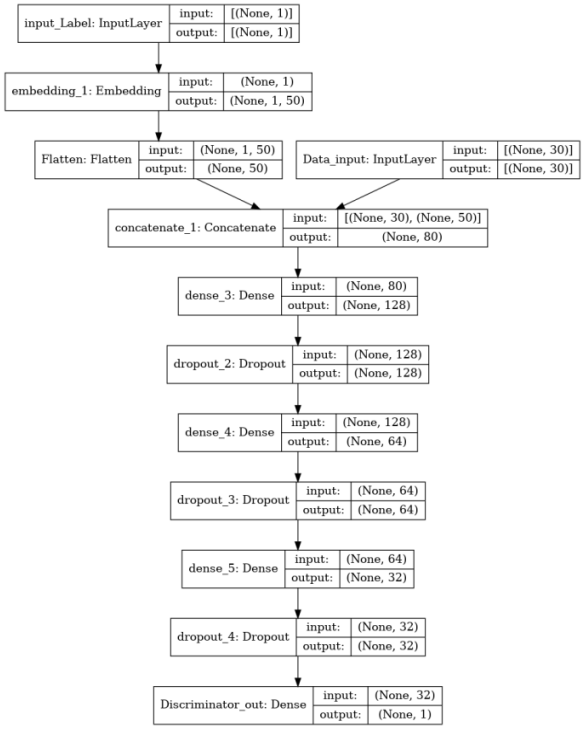


FIGURE 2. The discriminator network of Wasserstein GAN architecture.

the function. Conversely, it saturates the Generator. In other words, the Generator may end training if it is unable to catch with the Discriminator.

In order to solve the issue of saturation, a subtle variation of the conventional loss function is utilized. In this variation, the Generator maximizes the log of the Discriminator probabilities -$log\,(D\,(G\,(z)))$.

Figures 3 and 4 present the NS-GAN's Generator and Discriminator architectures, respectively.

This modification is aimed to frame the issue from a viewpoint, where the Generator tries to maximize the probability of instances being real, as a replacement of minimizing the probability of an instance being non-real. This process shuns the saturation of generator through a stable weight update mechanism.

Non-saturating GANs are alternatives to GANs such as, WGANs and f-GANs. The empirical work by [43] showed that the non-saturating model approximately minimizes the f-divergence KL, which the authors named softened reverse KL. The empirical study conducted by [33] confirms that softened reverse KL has a steeper slope in the left tail and slightly changes the behaviour of the right tail.

$$J^{(G)}(G) = -E_{z \sim pz} log D(G(z)) \quad (4)$$

Here, G is the model for probability distribution p(x).

D. LEAST SQUARE GAN

The study [44] argues that issues such as vanishing gradients may arise during training owing to loss of functions.

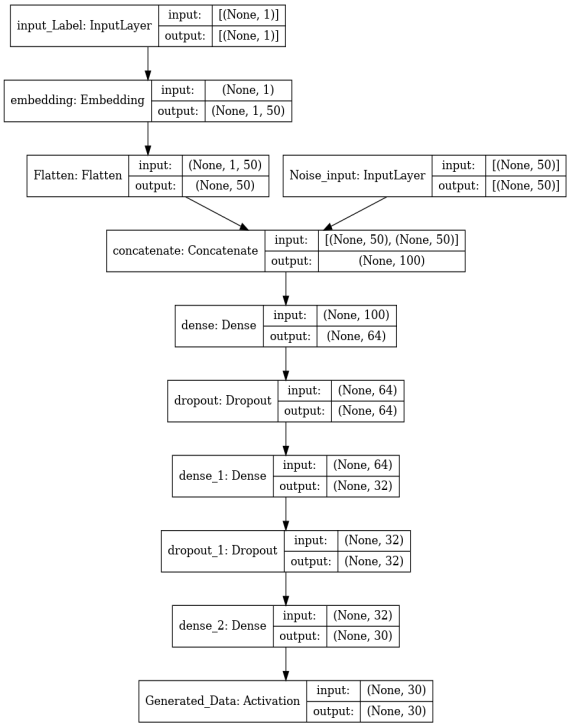


FIGURE 3. The generator network of non-saturating GAN architecture.

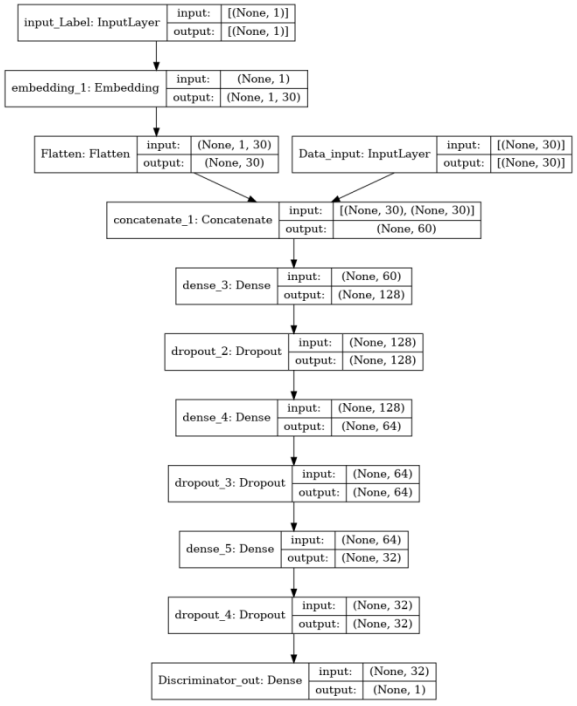


FIGURE 4. The discriminator network of non-saturating GAN architecture.

To address this challenge, they proposed the least square GANs (LS-GAN), which incorporates the least square loss function for the Discriminator [34]. Their empirical study confirmed that minimizing the objective function of the LS-GAN can minimize Pearson χ^2 divergence.

Figures 5 and 6 presents the Generator and Discriminator architectures of LS-GAN, respectively.

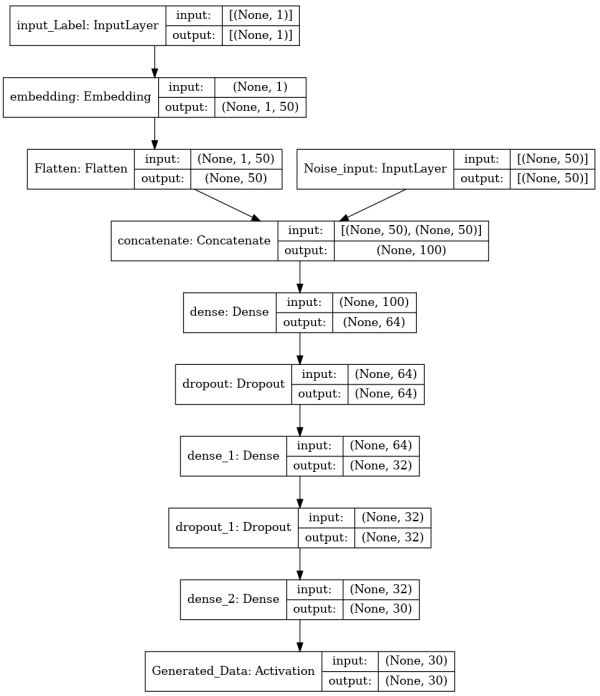


FIGURE 5. The generator network of least-square GAN architecture.

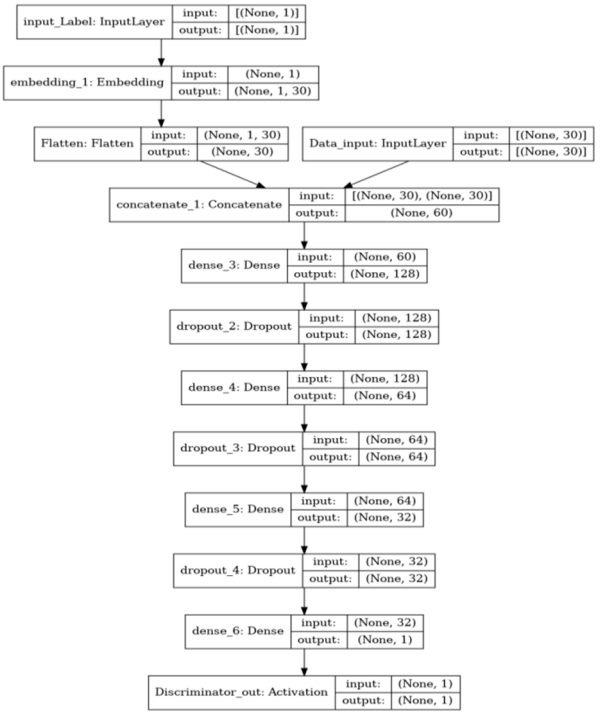


FIGURE 6. The discriminator network of least-square GAN architecture.

In GANs, the learning process is to train the Generator and the Discriminator at the same time. The objective of

Generator is to learn the distribution P_g over data "x". The sampling of input variables "z" is dome of the Generator from a uniform distribution $P_z(z)$. Moreover, the Generator guides the input variable "z" to data space $G(z; \ominus_g)$ via a differentiable network. At the same time, the Discriminator network is a classifier that tries to identify whether the instance is from the Generator or from real data.

Mathematical representation of loss functions of LS-GAN:

$$min_D V_{LSGAN}(D) = \frac{1}{2}E_{x\sim pdata(x)}[\left(D(x) - b\right)^2] + \frac{1}{2}E_{z\sim pz(Z)}[(D(G(z)) - a)^2]. \quad (5)$$

$$min_G V_{LSGAN}(G) = \frac{1}{2}E_{z\sim pz(Z)}[(D(G(z)) - c)^2] \quad (6)$$

where, G is the Generator, D is the Discriminator, "a" is the label for fake data, "b" are the labels for real data and "c" is the value that the Generator wants the Discriminator to perceive as fake data.

E. SYNTHETIC DATA GENERATION GAN

In their study, [45] the proposed synthetic data generation GAN, or SDG GAN, was used to generate artificial data to train a supervised classifier. Their empirical study showed that the SDG GAN can perform better than density-based oversampling models and improves dataset classification.

The Discriminator and Generator of SDG-GAN are feed-forward networks with MLP framework. In SDG-GAN feature matching (FM) loss is utilized instead of regular loss. The FM loss is used to improve the training process of GAN. Moreover, SDG-GAN is developed on conditional GANs. The SDG-GAN paper utilized FM method for training the Generator. The objective of FM method is to change the cost function of the Generator to reduce the difference between the features of the produced data and real data. The objective function of FM loss is given below:

$$FM = \|E_{x\sim pdata} f(x) - E_{z\sim P_z(z)} f(G(z))\|_2^2 \quad (7)$$

The f (x) in FM loss is the feature vector extracted by a layer in the Discriminator. The objective of FM is to address the instability of GANs via giving new objective to the Generator which averts it from over-training.

Additionally, the SDG-GAN paper utilized a conditional GAN (cGAN) structure for estimation of conditional distribution, p_y^x. This process allowed sampling of minority class label, $X_{new} = G(z, y)$.

The objective function of the SDG GAN is:

$$min_G max_D \|E_{z\sim pdata} f\left(x/y\right) - E_{z\sim p_z\left(z/y\right)} f(G(z))\|_2^2 + E_{x\sim pdata}[log\left(D\left(x/y\right)\right)] \quad (8)$$

Moreover, Figures 7 and 8 present the Generator and Discriminator architectures of SDG GAN.

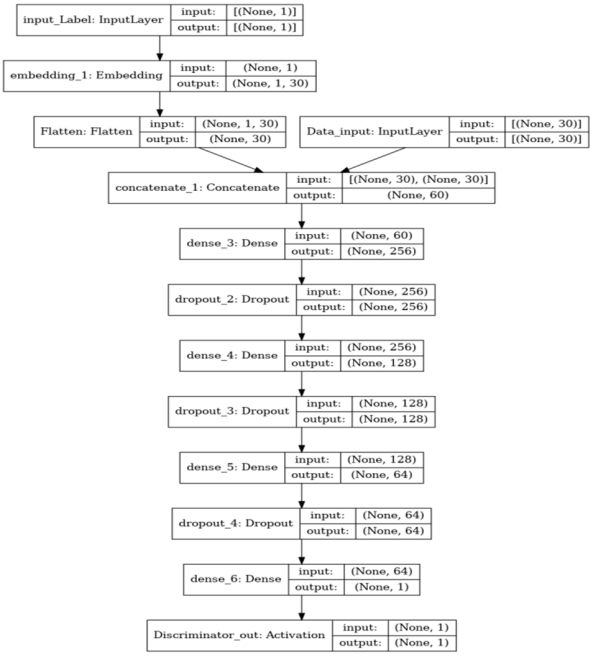


FIGURE 7. The discriminator network of SDG GAN architecture.

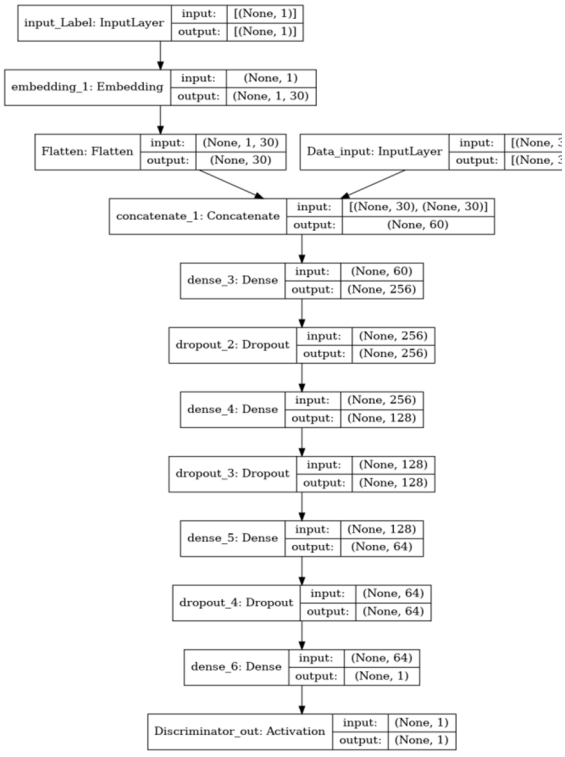


FIGURE 8. The discriminator network of SDG GAN architecture.

F. PROPOSED METHOD: NOVELTY K-CGAN

Our proposed algorithm, Novelty K-CGAN, comprises two main sub-networks: Generator and Discriminator.

The initial stage of training the K-CGAN involves generating Random Noise with batch size and latent dimension, where the batch size is the size of the input batch of data and

the latent dimension is the dimension of the noise. Afterwards the Generator generates fake data. The Generator of novelty K-CGAN transforms the random noise and labels into fake data. On the other hand, the Discriminator links both the real data and the generated synthetic data with their labels to shape the combined labels and combined data. Our proposed framework then prepares target labels to form a binary label that discriminates real data from fake data. The next phase of K-CGAN involves training the Discriminator. In this phase, the Discriminator is used to classify the combined labels and data to minimize the Discriminator loss, defined by the loss function between the predictions and true labels.

Moreover, the gradients of the loss with respect to the Discriminator's trainable weights were computed and used to update the parameters of the Discriminator. The Discriminator's loss was calculated in the next stage of training. The Discriminator loss is the binary cross-entropy between misleading labels and prediction. Furthermore, the gradients and update weights were calculated. This process is known as back-propagation, in which the gradients of the loss with respect to the weights are calculated. The optimizer was then utilized to update the weights, ultimately leading to the minimum possible loss. The next steps of training novelty K-CGAN involve the preparation of data and the start of the Generator training. In this training phase, the Generator is trained to transform the noise and real labels into fake data, and uses the Discriminator to classify these fake data samples. Moreover, the Generator aims to maximize the loss defined by the loss function between the misleading labels and the Discriminator's predictions of the fake data.

The Generator loss was calculated after the Discriminator loss. The Generator loss consists of two terms: binary cross-entropy between true labels and predicted labels by the Discriminator, and KL Divergence between the original data and fake data generated by the Generator. The Kullback-Leibler (KL) divergence is a measure of the difference between two probability distributions. This is used to optimize the parameters of the Discriminator model based on the difference between the original data distribution and the fake data distribution of the Generator model. The next stage of training the K-CGAN is referred to as backpropagation, where the gradients of the loss with respect to the weights are calculated. The optimizer was then utilized to update the weights, ultimately leading to the minimum possible loss. In the last step of training, tracking of the Discriminator and the Generator loss is performed using the Generator loss tracker and Discriminator loss tracker objects, returning these losses as a dictionary. The diagrammatical representation of our proposed model is given below:

Figure 9 shows the basis architecture of K-CGAN. We can see that the Discriminator is the binary cross-entropy between misleading labels and prediction. Whereas, the Generator loss is comprised of two terms: binary cross-entropy between true labels and predicted labels by the Discriminator, and KL Divergence between the original data and fake data generated by the Generator.

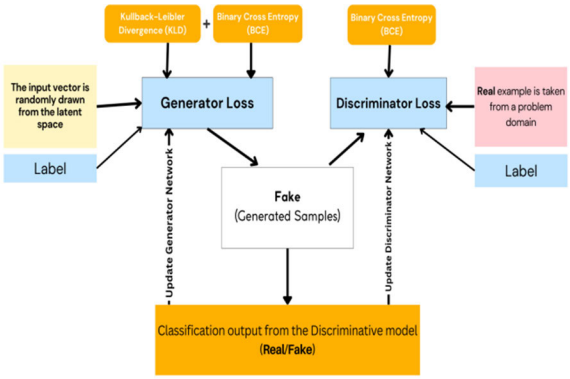


FIGURE 9. Basic architecture of K-CGAN approach.

G. NOVELTY K-CGAN LOSS FUNCTION

In this study, we defined a new novelty loss function that adds the KL divergence loss to ensure that both distributions are close to each other. Furthermore, the study performed hyperparameter tuning to determine the optimal weight for the KL divergence loss.

K-CGAN **Discriminator** Loss:

The Discriminator loss of experimental K-CGAN is Binary Cross Entropy.

$$\text{Discriminator Loss} = -\frac{1}{\text{output size}} \sum_{i=1}^{\text{output size}} \text{yi}.\log \hat{\text{y}}\text{i} + \left(1 - \hat{\text{yi}}\right).\log\left(1-\hat{\text{y}}\text{i}\right) \tag{9}$$

The Discriminator loss of K-CGAN has two parts: the binary cross-entropy loss for the real data and the binary cross-entropy from the generated data. The first part deals with the accuracy of the Discriminator to identify real data. The Discriminator takes in the original data and creates a prediction, which is compared to the true label. The objective is to lessen the loss term. Moreover, the second term, binary cross-entropy for generated data, measures the efficiency of the Discriminator network to discriminate real and generated data. The objective is to lessen the loss term.

Figure 10 shows the Discriminator network of K-CGAN.

K-CGAN **Generator** Loss:

The Generator loss of the experimental K-CGAN combines the binary cross-entropy and **KL divergence.**

$$\begin{aligned} &\textit{Generator Loss} \\ &= -\frac{1}{\textit{output size}}\left(1-\hat{yi}\right) \\ &\quad + \sum_{i=1}^{outputsize} yi.log\hat{y}i + \left(1-\hat{yi}\right).log\sum pi(x)log(\frac{pi(x)}{qi(x)}) \end{aligned} \tag{10}$$

Like the Discriminator loss, the Generator loss has two parts: the binary cross-entropy loss and the KL divergence. The binary cross-entropy loss measures the ability of the Generator to trick the Discriminator. The purpose of the Generator is to take in a noise vector and generate artificial data. The binary cross-entropy loss compares the resulting prediction

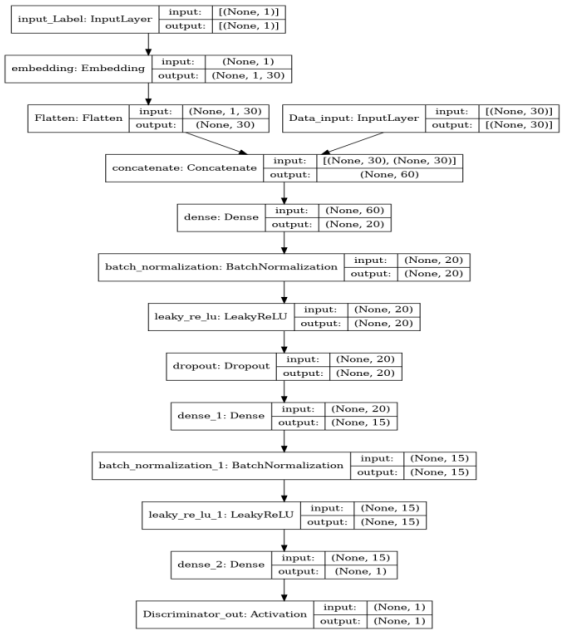


FIGURE 10. The discriminator network of K-CGAN GAN architecture.

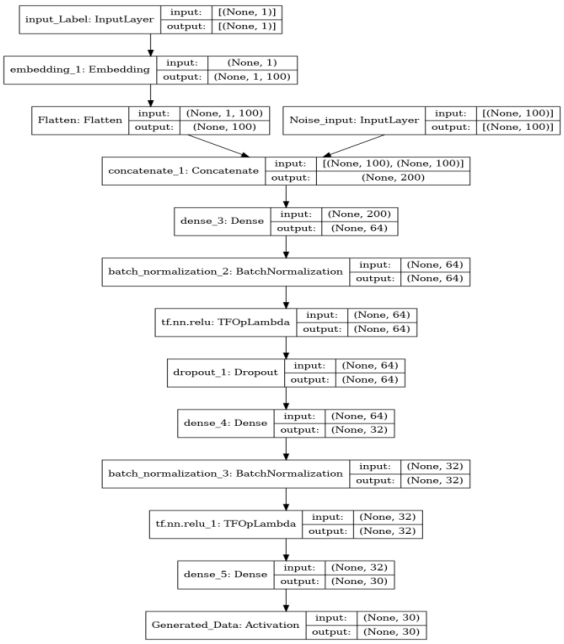


FIGURE 11. The generator network of K-CGAN GAN architecture.

with the real label. The objective here is to maximize the loss term. In other words, the Generator should produce artificial data that should be difficult for the Discriminator to distinguish from the original data.

On the other hand, the KL divergence term in the Generator loss computes the difference between the distributions of the generated data and real data. The objective is to lessen the loss term. By minimizing the KL divergence, the Generator learns to generate artificial data resembling real data.

Binary Cross Entropy and KL Divergence impacts training by providing a balance between generating high-quality samples and ensuring that the generated samples match the target distribution.

Figure 11 show the Generator network of proposed K-CGAN method.

H. ADVERSARIAL TRAINING OF NOVELTY K-CGAN

In this section, we explained the framework of the proposed model.

The Generator loss consists of two terms: (a) binary cross-entropy between true labels and predicted labels by the Discriminator, and, (b) KL divergence between the original data and false data generated by the Generator.

Binary cross-entropy (BCE) is the negative average of the log of corrected predicted probabilities. It compares the predicted probabilities to actual class output. Moreover, based on the distance from the expected values, binary cross-entropy measures the score that penalizes the probabilities. By doing so, it computes the distance from actual value. The mathematical representation of BCE is given in below equation:

$$BCE = -\left({}^{1}/_{N}\right)\left(y_{predicted}\right) + \sum y\_true\, log\,(1 - y_{true}) * log(1 - y\_predicted) \quad (11)$$

In the above equation, "y" is the label and "$p(y)$" is the predicted probability of point being positive for each "output size" points. Moreover, for each "y", it adds "$log(y\_predicted)$" to the loss- chances of the point being true. On the other hand, $\log(1 - y\_predicted)$ is the chance of being negative, for every negative point "$y{=}0$".

The mathematical representation of KL divergence is shown in the below equation:

$$KL\left(original_{data\|fake_{data}}\right) = \sum original_{data} * log({}^{Original\_data}/_{fake\_data}) \quad (12)$$

In the above KL divergence equation, there are two probability distributions, "Original_data" and "Fake_data" distributions.

In our proposed novelty K-CGAN approach, the KL divergence is utilized to minimize the difference between the distribution of the generated data using GAN and the distribution of the real data. By doing so, the K-CGAN is encouraged to generated data that closely resemble the traits of the real data

The study utilized the Wisconsin Breast Cancer Dataset. The following steps were taken to train the K-CGAN network using the novelty Loss function.

The K-CGAN training procedures can be briefly presented as follows:

IV. EXPERIMENTAL SETUP

This study utilized multiple GAN variants for data augmentation to demonstrate how classification methods work with small tabular medical datasets. The study proposed a K-CGAN technique for data augmentation and compared it with other state-of-the-art GAN frameworks (LS-GAN, WGAN, NS-GAN, and SDGs-GAN).

Algorithm 1 Training Procedure Using K-CGAN

Generate Random noise (noiseD)
With Shape (batch_size, 2:self.latent_dim),
(batch_size)= size of the Input Batch of Data
(latent_dim) = dimension of the Noise.
Generate fake data
#using the Generator to Transform the Random
noise(noise) and (real_labels) Into Fake Data.
Data preparation for the Discriminator
Training phase
#The real data (real_data) and generated fake data
(generated_data) Along With Their labels
(real_labels) to form (combined_data) and
(combined_labels)
Target labels are prepared
Binary labels (labels) are formed
Use discriminator to classify combined_data and
combined_labels
The goal is minimize the loss (d_loss)
Prepare data for generator training
Generator training during training
Random noise (noiseG) is generated with
(batch_size,Self.latent_dim) to train the
generator
Uses generator
#transform noise (noiseG) and labels
(real_labels) into (fake_data)
Use discriminator to classify these fake data.
#the generator aims to maximize the loss (g_loss)

#the generator loss is equal to BCE + KL
Calculations
#Backpropagation process: gradients of the loss with
Respect to the weights are calculated
Goal is minimum Loss
Tracking step
(g_loss) and (gen_loss) tracking using
(gen_loss_tracker) and (disc_loss_tracker)
#Returns these losses as dictionary

A. DESCRIPTION ABOUT THE DATASET

In this study, we used the well-known Breast Cancer Wisconsin (Diagnostic) Data Set (WBCD), which can be easily accessed online [46]. This dataset was comprehensively used for the model-based classification. The Wisconsin Breast Cancer dataset was used to evaluate the performance of the algorithms. The breast cancer dataset was selected to test the strategies at multiple levels of complexity. Moreover, this study used a breast cancer dataset to verify the accuracy of the algorithms with augmentation techniques and authenticate the proposed GAN-based method.

The dataset contains 31 columns, and the column number is the first column. The dataset comprised 569 patients with breast tumours, of which 357 were negative (0) and 212 were positive (1).These tumour cases were characterized by 32 features, including area, texture, radius, perimeter, concavity, compactness, smoothness, fractal dimension, and symmetry. Moreover, each of the 32 features was characterized by the: worst value, mean, and standard error. The worst value represents the outlier in measurements, that is, values that are not in the medically specified range. Additionally, all instances are identified by a pseudonym and labelled as benign or malignant. Table 1 summarizes the breast cancer datasets used in this study. Figure 12 shows the class distribution into the minority and majority classes in the breast cancer dataset.

TABLE 1. Basic information about dataset.

Dataset	No. of Attributes	No. of Samples	Class Distribution (Benign/Malignant)
WBCD	32	569	357/212

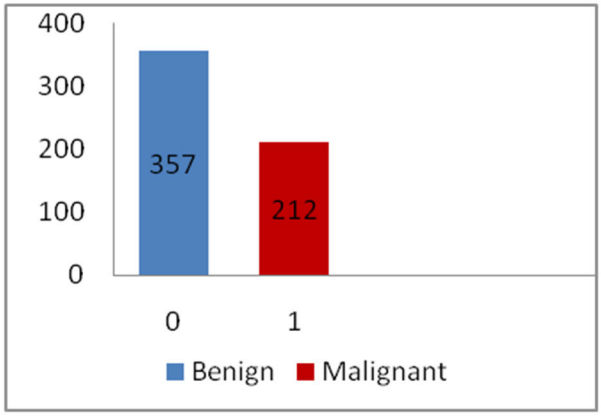


FIGURE 12. Class distribution of breast cancer dataset. This figure indicates the WDBC classification into benign and malignant cases.

B. PERFORMANCE MEASUREMENT

To quantitatively analyze the performance of our proposed method in the breast cancer domain, we used the following performance metrics: precision, F1 Score, recall, and accuracy. These performance metrics are widely used for classification problems. In our cancer detection model, we calculated the following metrics:

$$Accuracy = \frac{TP + TN}{N} \tag{13}$$

$$F1\ Score = \frac{2 \times precision \times recall}{precision + recall} \tag{14}$$

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

The values of precision can vary between 0 and 1.

Here, N is the sum of the samples and TP is the number of fault samples S, which are classified correctly. On the other hand, FP is the number of other category samples predicted incorrectly to be in sample S. FN is the number of current fault samples classified incorrectly as other categories.

Moreover, we used classifiers with their default setting values to train on the original dataset and then used it to calculate the performance metrics on the synthetic data generated by the GAN model containing equal class distribution of 357 benign and 357 malignant samples. A high value indicates that the synthetic data generated by the GAN model accurately represents the original data.

C. MACHINE LEARNING ALGORITHMS

Machine learning methods employ an inference principle known as induction, which indicates that conclusions can be drawn from a small set of examples. Supervised learning is typically used for induction. In supervised learning, a dataset of inputs and required outputs is employed to signify the modelled problem. For new inputs, the machine learning algorithm uses this information representation extracted from the examples to create the output. For instance, if there are n examples in an equation (Xi;Yi), where Xi signifies the input and Yi indicates the output. Here, we can view the obtained classifier as a function f. The function obtains input X and returns output Y [47].

This section presents a general review of machine learning methods used in this study. All techniques used in this study have a distinctive approach, and we selected these techniques as promising examples from different types of learning.

The Random Forest (RF) is an ensemble classification technique in which each model comprises decision trees that are trained randomly to lessen the association. Random Forest is easy to train and provides precise output predictions. It also makes rational estimates of the feature importance. When compared with a Decision Tree, Random Forest can be more difficult to implement owing to the greater number of parameters [48].

The logistic Regression classifier is a statistical model that explicates the chances of the occurrence of a class by fitting a logistic curve to the dataset [49]. Logistic Regression classifiers are also known as logit, maximumentropy, and logistic model classifiers. These models are extensively used in statistics.

The Extreme Gradient Boosting (XGBoost) classifier is a learning method that combines the predictions of several weak models to generate stronger predictions. XGBoost is one of the most admired and commonly used machine learning methods because of its capability to achieve excellent performance in multiple tasks, such as regression and classification [50].

The KNN is a basic yet vital classification method used in machine learning. It can be used in real-life examples, as it does not make underlying conjectures about data distribution [51].

In addition to the above machine learning algorithms, this study also used the Multilayer Perceptron (MLP) algorithm. MLPs have the ability to train on a set of input-output pairs and learn to model the dependencies or correlations of the inputs and outputs [52]. The training of the MLP included adjusting the parameters to reduce errors.

D. FEATURES AND TARGET COLUMN

We used the feature column to prepare the dataset for training. In addition, the target 'Diagnosis' column is used for the prediction label.

1) FEATURES COLUMNS

We applied normalization to scale feature values between 0 and 1. Min-MaxScaler is a pre-processing technique used in machine learning to scale the features of a dataset to a specific range. The purpose of scaling the features is to ensure that all features have a similar impact on the model and to prevent certain features from dominating the model because of their large values.

The method can be summarized as follows:

Algorithm 2 Feature Selection Procedures

1: Epochs: 10000
2: Batch_size = 64
3 #How Many Data in one Batch
4:HP_NOISE = 100
5: #Lengh of Noise Vector
6: HP_DROPOUT = 0.2
7: #Dropout to Be Used in the Neural Network
8: HP_WEIGHTS_INIT = 'glorot_uniform'
9: #Weight Initialization
10: HP_DISCRIMINATOR_LAYERS = '2, -20, 10'
11: # Defined three possibilities of hidden layers.
12: HP_GENERATOR_LAYERS = '2, -128, 64'
13: #Here we have defined three possibilities of hidden layers for the generator model.
14: SAMPLES_COUNT = 400
15: #How many fake samples to be generated while calculating F1 Score
16: EARLY_STOPPER_PATIENCE = 50
17: #Stop training if accuracy doesn't improved for continuous n epochs.

V. RESULTS

In this study, five classifiers were used. We synthetically generated a dataset using a saved Generator model. After using the trained classifier, we calculated the F1 Score on the synthetic dataset and compared it with that of the original dataset. Furthermore, we checked both Generator and Discriminator losses. The results of different GAN variants in terms of Discriminator and the Generator losses are presented in this section.

A. TRAINING AND TESTING

For our proposed method, we selected multiple hyperparameters to achieve the maximum performance. After checking

multiple hyperparameters, we chose the hyperparameters below to achieve better results.

Table 2 and Table 3 list our selected settings, where Table 2 presents the hyperparameter settings for Generator and Table 3 presents hyperparameter settings for Discriminator.

TABLE 2. This table shows the generator hyperparameter settings.

Parameter	Quantity
Learning rate	0.0001
Hidden Layer Optimizer	Relu
Output Optimizer	Adam
Loss Function	Trained Discriminator Loss + KL Divergence
Hidden Layers	2, -128, 64
Dropout	0.1
Random Noise Vector	100
Kernel Initializer	glorot-uniform
Kernel Regularizer	L2 method
Total Learning Parameter	36, 837

TABLE 3. This table shows the discriminator hyperparameter settings At10000 epochs.

Parameter	Quantity
Learning Rate	0.0001
Hidden LayerOptimizer	LeakyRelu
Output Optimizer	Adam
Loss Function	Binary Cross Entropy
Hidden Layers	2, -20, 10
Dropout	0.1
Kernel Regularizer	L2 method

We used a learning rate of 0.0001 and hidden layers of 2, -128, 64, and 2, -20, 10 for the Generator and Discriminator, respectively, as shown in Table 2 and Table 3. We set the dropout ratio at 0.1 on the Generators and Discriminators hidden layers. Furthermore, the random noise vector for the Generator is 100. For training, we selected the Adam output optimizer. Similarly, in terms of the activation function, we used a learning rate of 0.0001 and hidden layers of 2, -128, 64, and 2, -20, 10 for the Generator and Discriminator, respectively, as shown in Table 2 and Table 3. We set the dropout ratio at 0.1 on the Generators and Discriminators hidden layers. Furthermore, the random noise vector for the Generator is 100. For training, we selected the Adam output optimizer. Similarly, in terms of the activation function, we used a rectified linear unit (ReLU) as the hidden layer optimizer. Moreover, for the Discriminator, we used binary cross-entropy loss functions; for the Generator, we used the Trained Discriminator Loss + KL Divergence as the loss function.

B. EXPERIMENTAL K-CGAN

Figures 13 and 14 show the Loss of K-CGAN (Generator loss and Discriminator loss).

FIGURE 13. EXPERIMENTAL K-CGAN's generator loss.

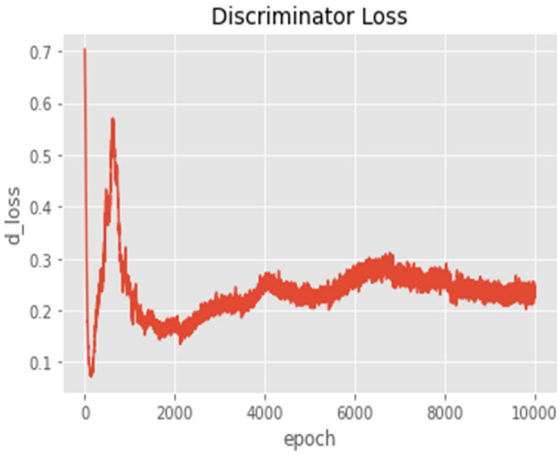


FIGURE 14. EXPERIMENTAL K-CGAN's discriminator loss.

Initially, the Discriminator loss was unstable;however,after a few epochs, the Discriminator loss was stable and oscillating.

The Generator and Discriminator loss shows that our K-CGAN' Discriminator loss is very smooth and oscillates around a very small number between 0.2 and 0.3.

In addition, the Generator loss of K-CGAN was unstable at the initial stage but after few epochs the Generator loss becomes smooth and stable. The experimental observations suggest that K-CGAN method performed very well on breast cancer dataset.

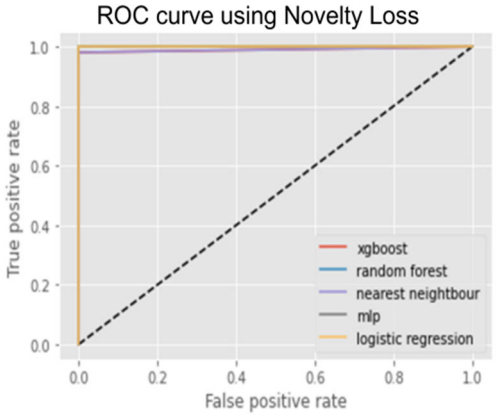


FIGURE 15. This figure shows the ROC curve using the K-CGAN.

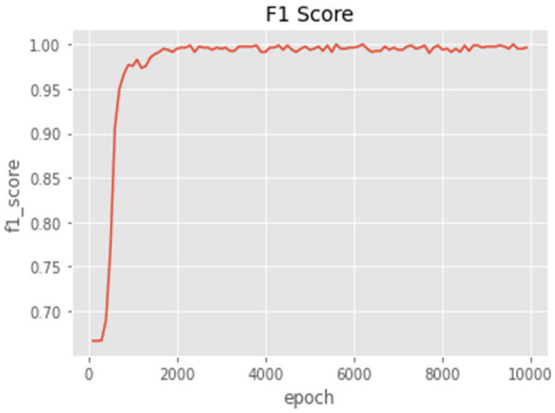


FIGURE 16. This figure shows the F1 Score using the K-CGAN.

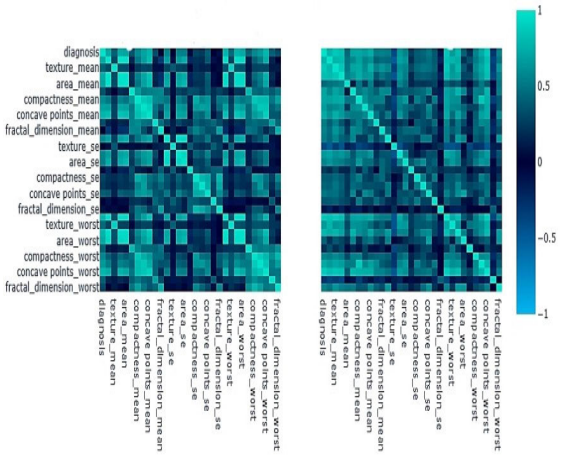


FIGURE 17. Numerical correlation (real data) and numerical correlation using K-CGAN (synthetic data).

TABLE 4. Classification performance of K-CGAN.

Model	Precision	Recall	F1 Score	Accuracy
XGBoost	0.982801	1.0	0.991326	0.99125
Random Forest	0.977995	1.0	0.988875	0.98875
Nearest Neighbor	0.982801	1.0	0.991326	0.99125
MLP	0.985222	1.0	0.992556	0.99250
Logistic Regression	0.977995	1.0	0.988875	0.98875

On the other hand, Figures 15 and 16 show the ROC curve using Novelty Loss and F1 Score of Novel K-CGAN method, respectively. We can see that the F1 Score improved very smoothly, and after 2000 epochs, it was 0.99+. After every 100 epochs, we calculated the F1 Score and plotted its results of the F1 Score. In addition, the total epoch run was 10,000 epochs.

On the other hand, Figure 17 presents the Numerical Correlation (Real Data) and Numerical Correlation (Synthetic Data) using K-CGAN.

In addition, the Table 4 shows the classification performance of K-CGAN method. The performance was evaluated on the basis of Recall, Precision, Accuracy values and F1-Scores. The values illustrates that the K-CGAN method performed well than other GAN methods.

Moreover, the Table 9 presents a detailed comparison of K-CGAN method with five well known GAN variants. When compared with other methods, the experimental K-CGAN performed the best with the Breast Cancer Dataset, and the NS GAN performed the 2nd best performing one.

C. LEAST SQUARE GAN (LS GAN)

The Generator loss of Least Squares Generative Adversarial Network (LS-GAN) was highly unstable during the initial phase of training. Over 1700 epochs, the training stability gradually improved but still remained far from smooth. Figure 18 illustrates this behavior for the Generator loss over time. Notably, the loss fluctuated between 0.08 and 0.18 before gradually stabilizing. These results demonstrate the difficulty of training LS-GAN and suggest that alternative approaches may be necessary for more robust performance. Nevertheless, the study highlights the potential of generative adversarial networks in producing state-of-the-art results.

In contrast, the Discriminator loss of LS-GAN was remarkably unstable at the initial phase of training. Yet, after 1500 epochs, it became considerably stable with some oscillations. However, there were multiple patches of instability between 4500 to 7500 epochs as seen in Figure 19. This means that overall, the Discriminator loss of LS-GAN shows stability in few areas but not consistently throughout its range, oscillating between 0.11 and 0.13. These results suggest that while LS-GAN shows promise as a generative learning technique, there is still much work to be done in terms of improving the stability of its training process. As such, further research into improving the stability of LS-GAN should be conducted in order to maximize its potential as a powerful generative deep learning technique.

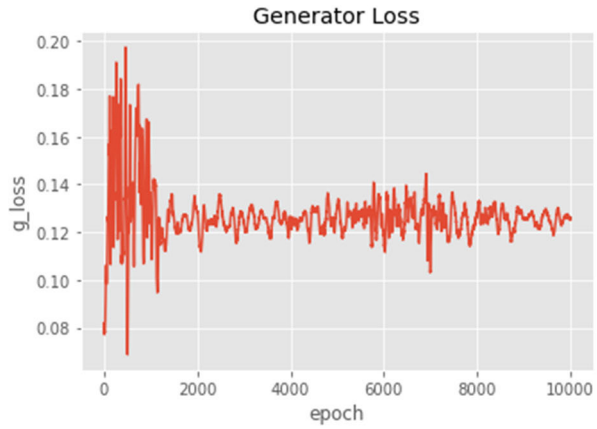


FIGURE 18. The generator loss of LS GAN.

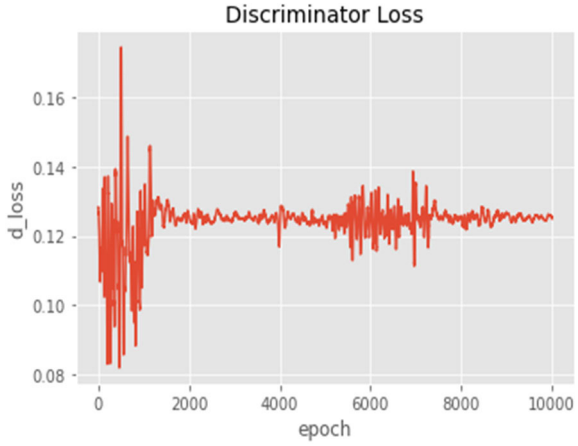


FIGURE 19. The discriminator loss of LS GAN.

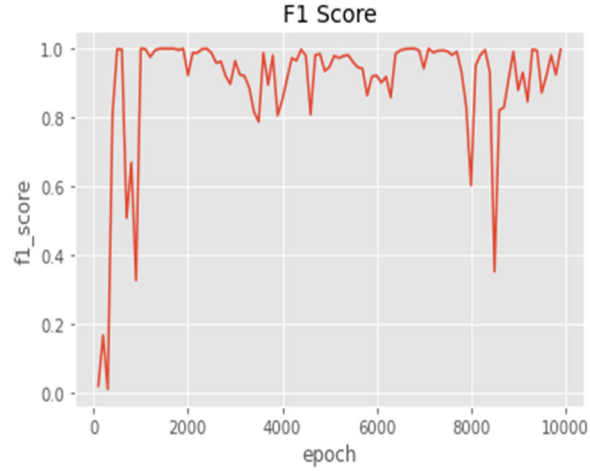


FIGURE 20. This figure shows the F1 Score using the least square GAN.

Figures 18 and 19 present the Generator loss and Discriminator losses of LS GAN, respectively. On the other hand, Figure 20 shows the F1 score using LS GAN.

To test the significance of the LS-GAN, we compared the F1 Score of LS-GAN and our K-CGAN model on the breast cancer dataset. The F1 Score using LS-GAN, as shown in Figure 20 indicates that the LS-GAN method cannot perform well on breast cancer dataset. However, the F1 Score was more stable than was expected.

In the experimental study, we noticed that sometimes the Generator generated synthetic data that were similar to the original dataset. Hence, from empirical evidence, the least square GAN did not perform very well. Figure 21 shows the ROC curve using LS GAN method. Moreover, Figure 22 presents the Numerical Correlation (Real Data) and Numerical Correlation (Synthetic Data) using LS GAN.

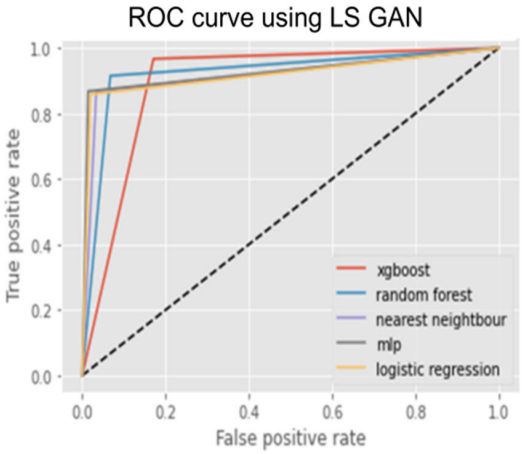


FIGURE 21. This figure shows the ROC curve using the least square GAN.

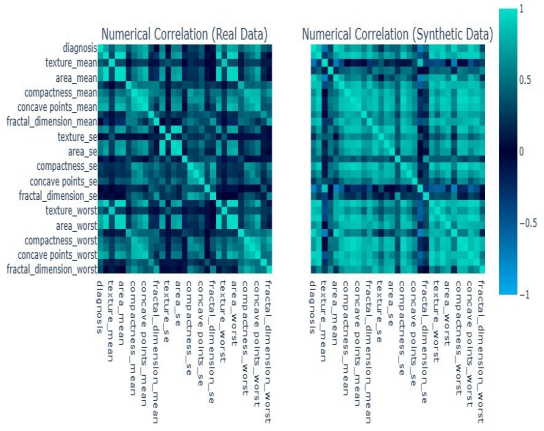


FIGURE 22. Numerical correlation (real data) and numerical correlation (synthetic data) using LS GAN.

Moreover, Table 5 presents the classification performance of LS-GAN. We evaluated the performance of LS-GAN using five classification methods. When compared with our proposed K-CGAN, the LS-GAN performed poorly.

D. WASSERSTEIN GAN

Our study used WGAN to check its applicability to the breast cancer dataset. It was observed that the WGAN could not

TABLE 5. Classification performance of least square GAN.

Model	Precision	Recall	F1 Score	Accuracy
XGBoost	0.770677	**0.966981**	0.857741	0.880491
Random Forest	0.889908	0.915094	0.902326	**0.926186**
Nearest Neighbor	**0.938776**	0.867925	0.901961	0.929701
MLP	**0.973545**	0.867925	0.917706	0.942003
Logistic Regression	**0.962963**	0.858491	0.907731	0.934974

learn at approximately 8000 epochs, and its performance declined after 8000 epochs. Furthermore, the Discriminator loss did not perform well, and remained at 0.00.

Figure 23 shows the Generator loss of WGAN. We noted that the Generator loss of LS-GAN was extremely unstable at oscillated between 0.08 and 0.18. The Generator loss suggests that WGAN performs poorly on breast cancer dataset.

FIGURE 23. The generator loss of WGAN.

Figure 24 presents the Discriminator loss of WGAN. It was observed that the Discriminator loss of WGAN did not perform well. Moreover, Figures 25 and 26 present the ROC curve and the F1 Score using the WGAN method, respectively. Initially, the F1-Score remained low and considerably stable. The F1-Score remained low till 4000 epochs, but after 4000 epochs it became extremely unstable and high. From the experimental observations we can say that the WGAN method is not a suitable method for the breast cancer dataset used in our study. Overall, the F1-Score was very low at the initial stage but at the later stage it became high.

Further, Figure 27 shows the Numerical Correlation (Real Data) and Numerical Correlation (Synthetic Data) using WGAN. The Table 6 shows the classification performance of WGAN method. When compared with the proposed K-CGAN method, the WGAN performed poorly on the basis

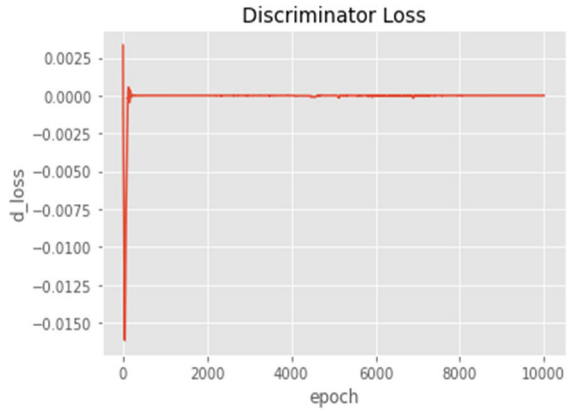


FIGURE 24. The discriminator loss of WGAN.

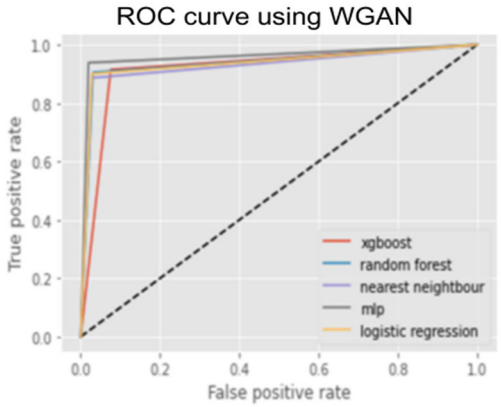


FIGURE 25. This figure shows the ROC curve using the WGAN.

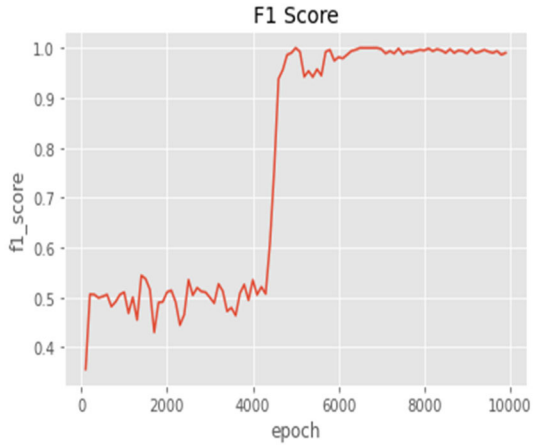


FIGURE 26. This figure shows the F1 Score using the WGAN.

of recall, accuracy, precision values and F1-score. For that reason, we can say that WGAN method do not perform well on breast cancer dataset.

E. NON-SATURATING GANs

In another study, [33] argued that when compared with the saturating training schemes related to the Jensen-Shannon

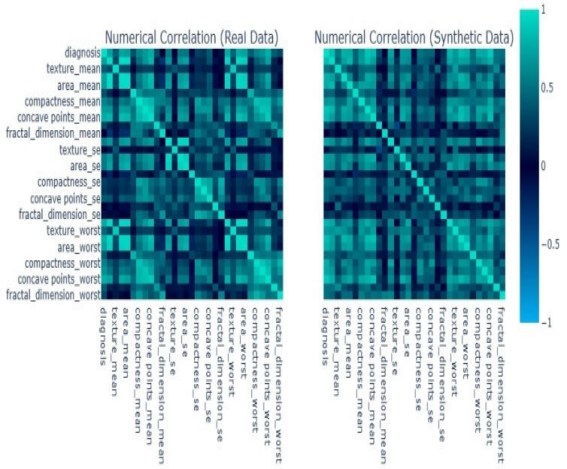


FIGURE 27. Numerical correlation (real data) and numerical correlation (synthetic data) using WGAN.

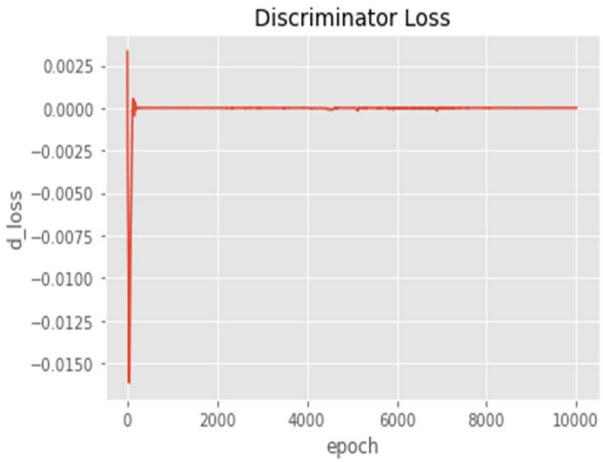


FIGURE 28. The discriminator loss using NS-GAN.

TABLE 6. Classification performance of wasserstein GAN.

Model	Precision	Recall	F1 Score	Accuracy
XGBoost	0.877828	0.915094	0.896074	0.920914
Random Forest	0.945813	0.905660	0.925301	0.945518
Nearest Neighbor	0.949495	0.886792	0.917073	0.940248
MLP	0.966019	0.938679	0.952153	0.964851
Logistic Regression	0.950219	0.900943	0.924939	0.945518

FIGURE 29. The generator loss of non-saturating GAN.

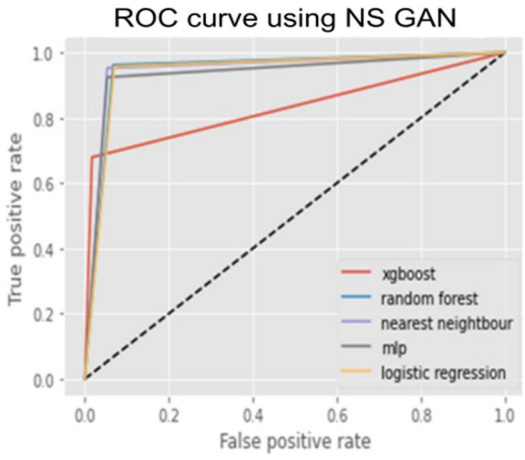


FIGURE 30. The ROC curve using NS-GAN.

divergence, the non-saturating schemes provide much better gradients at the early stage of training. To verify the performance of NS-GANs, we used them in our study on the breast cancer dataset.

Moreover, the Figure 28 shows the Discriminator loss using NS-GAN. It was observed that the NS-GAN performed poorly in terms of the Discriminator loss.

Figure 29 shows the Generator loss of NS-GAN. We noted that the Generator loss of NS-GAN remained highly unstable in the initial phase of training. After 1700 epochs the Generator loss recovered considerably. However, the loss was not smooth and remained unstable from time to time. Compared to our proposed –K-CGAN method, the Generator of NS-GAN remained unstable as it oscillates around a high number.

Figures 30 and 31 shows the ROC curve and the F1 score using NS-GAN. The results confirm that the F1 Score using NS-GANs is more than 80% but is neither still nor stable.

We observed that both the losses were stable after 1000 epochs. Moreover, the F1 Score was greater than 80%, but it could be more stable. Figure 31 shows that the F1 Score is approximately 0.85 at times, but usually reaches 1. Nevertheless, the Non-Saturating GAN performed reasonably well with the breast cancer dataset compared to the other GANs used in our experiment. However, the K-CGAN attained better and more stable results than the NS-GAN.

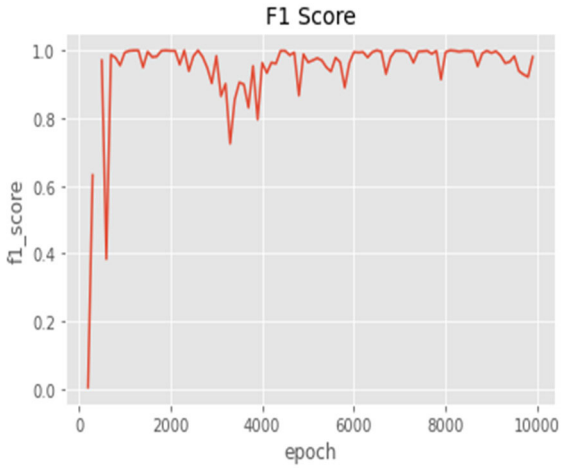


FIGURE 31. This figure shows the F1 Score using the non-saturating GAN.

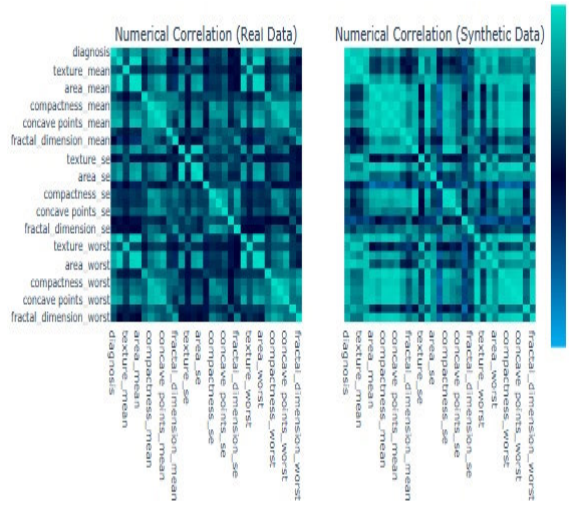


FIGURE 32. Numerical correlation (real data) and numerical correlation (synthetic data) using NS-GAN.

In addition, Figure 32 shows the numerical correlation (real and synthetic data) using LS-GAN.

In addition, Table 7 shows the classification performance of NS-GAN on the basis of Recall, Precision, Accuracy and F1 score.

F. SYNTHETIC DATA GENERATION GANs

Our empirical study shows that the SDG GAN model performs poorly for breast cancer datasets.

Figure 33 shows the Discriminator loss of SDG-GAN. We noted that the Generator loss of SDG-GAN was extremely unstable throughout the experiment. The loss oscillates between 0.04 and 4. Our experimental study confirmed that the SDG-GAN performs poorly on breast cancer dataset.

Figure 34 presents the Generator loss of SDG-GAN. It was observed that the Generator loss of SDG-GAN was stable at initial phase. However, after 8000 epochs, we observed

TABLE 7. Classification performance of non-saturating GAN.

Model	Precision	Recall	F1 Score	Accuracy
XGBoost	0.960000	0.679245	0.795580	0.869947
Random Forest	0.894737	0.962264	0.927273	0.943761
Nearest Neighbor	0.914027	0.952830	0.933025	0.949033
MLP	0.911628	0.924528	0.918033	0.988489
Logistic Regression	0.890351	0.957547	0.922727	0.940246

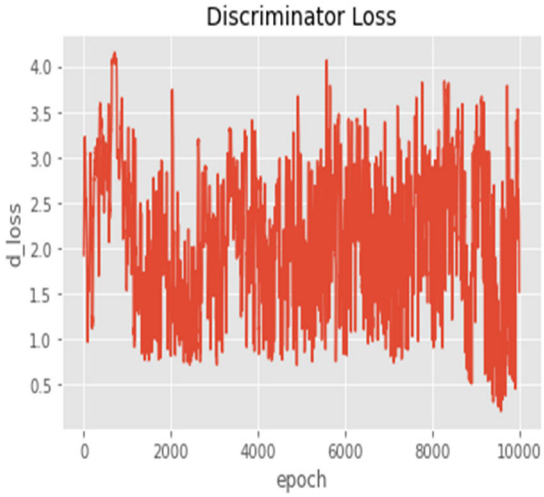


FIGURE 33. The discriminator loss of SDG GAN.

instability in the Generator loss and oscillated between 1 and 6. These observations confirmed that the SDG-GAN do not perform well on the breast cancer dataset.

Moreover, Figure 35 shows the ROC curve using SDG-GAN. Figure 36 shows the F1-Score using SDG-GAN. It was observed that there are many fluctuations in the F1 Score. The empirical findings suggest that F1-Score using SDG-GAN is highly unstable at every epoch.

Moreover, Figure 37 presents the numerical correlations (real and synthetic data) using SDG GAN.

In addition, the Table 8 presents the classification performance of SDG-GAN. The performance was evaluated on the basis of Recall, Precision, Accuracy values and F1-Score. The values confirmed that SDG-GAN performed poor when compared to K-CGAN and LS-GAN.

VI. DISCUSSION

This section presents a detailed discussion of the performance of our proposed method, K-CGAN. Our proposed method is based on a conditional GAN (cGAN) framework with the custom loss function of a generator. The study utilized the Kullback-Leibler divergence. Hence, we named our

FIGURE 34. The generator loss of SDG GAN.

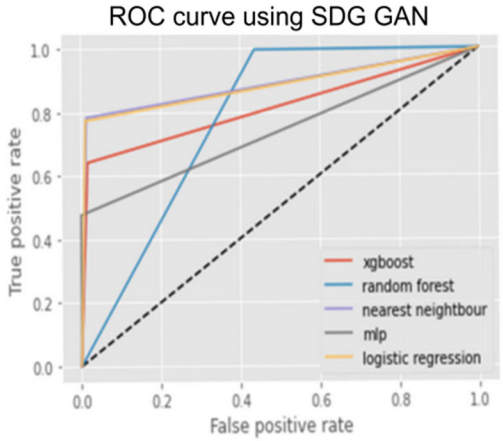


FIGURE 35. This figure shows the ROC curve using SDG GAN.

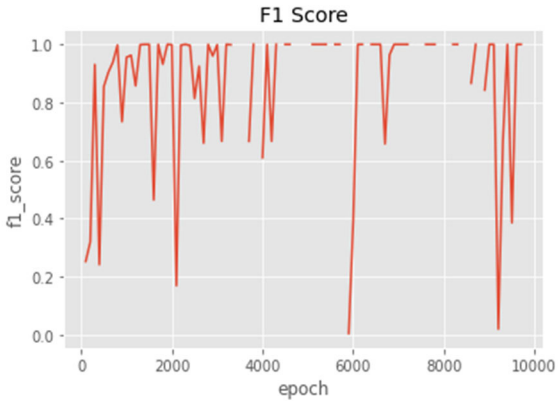


FIGURE 36. This figure shows the F1 Score using the SDG GAN.

proposed method Kullback-Leibler divergence Conditional GAN (K-CGAN) composed of two sub-networks: the Generator and the Discriminator. As mentioned in the previous

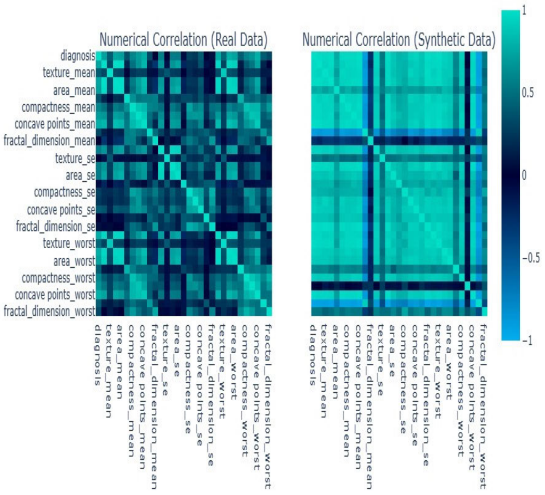


FIGURE 37. Numerical correlation (real data) and numerical correlation (synthetic data) using SDG GAN.

sections, we used K-CGAN to generate synthetic breast cancer data to improve the classification of malignant cases. The aim was to generate enough synthetic malignant cases to balance the original dataset. Moreover, the study compared our proposed method with other GAN variants.

One of the aims of this study was to generate synthetic data to avoid the need for large and complex datasets. In this way, the study was intended to deal with the issue of limited training data in the breast cancer domain. In this regard, the study implemented the WDBC dataset. Moreover, we synthetically generated a dataset utilizing a saved generator. After that, we calculated precision, recall, accuracy, and F1-score values on the generated dataset.

In this study, we analyzed the performance of K-CGAN on the breast cancer dataset and compared it with other LS-GAN, NS-GAN, WGAN, and SDG-GAN. The classification performance of K-CGAN was evaluated based on four performance metrics - Precision, Recall, F1 Score, and Accuracy.

To compare the performance of these GANs, we examined the results in Tables 4, 5, 6, 7, and 8, which display the classification performance of each model in terms of precision, recall, F1 score, and accuracy. Table 4 shows the performance of K-CGAN when applied to the breast cancer dataset. Regarding K-CGAN, we can see its classification performance is impressive, with high precision, recall, F1 score, and accuracy values across all models. XGBoost, Random Forest, Nearest Neighbor, MLP, and Logistic Regression all show high precision, recall, F1 score, and accuracy levels, each exceeding 97%. Specifically for the MLP model, K-CGAN has obtained almost perfect scores with a precision of 0.985222, recall of 1.0, F1 Score of 0.992556, and accuracy of 0.99250. These high scores demonstrate the efficacy of K-CGAN in detecting and diagnosing breast cancer with high accuracy. Based on these results, it is clear that K-CGAN is a powerful technique for breast cancer classification, performing consistently with all models.

TABLE 8. Classification performance of SDG GAN.

Model	Precision	Recall	F1 Score	Accuracy
XGBoost	0.957747	0.641509	0.768362	0.855888
Random Forest	0.574932	0.995283	0.728843	0.724077
Nearest Neighbor	0.970780	0.783019	0.868841	0.910369
MLP	1.000000	0.476415	0.645367	0.804921
Logistic Regression	0.976190	0.773585	0.863158	0.908612

On the other hand, compared to the performance of K-CGAN, LS GAN, WGAN, NS GAN, and SDG GAN do not demonstrate the same level of success in breast cancer classification based on their respective classification performance tables.

On the other hand, Table 5 shows the findings of LS GAN. However, the results show lower scores compared to K-CGAN, indicating the inferior performance of LS-GAN. For example, LS-GAN has achieved a precision score of 0.770677 for the XGBoost model, whereas K-CGAN has achieved a score of 0.982801 for the same model. Similarly, LS-GAN has obtained a recall score of 0.966981 for XGBoost, while K-CGAN has achieved a perfect recall score of 1.0. These results indicate that LS-GAN is less accurate in detecting and diagnosing breast cancer compared to K-CGAN. Moreover, the XGBoost model for LS GAN reached a recall value of 96.6%, the precision, f1-score and accuracy values were below 90%, indicating the model's inability to accurately classify breast cancer samples.

Similarly, Table 6 shows the performance of WGAN when applied on breast cancer dataset. Although the performance of the WGAN's MLP model was high with a precision of 96.6% and an accuracy of 96.4%, the remaining models' values were less impressive, and the WGAN's overall performance was poorer to K-CGAN.

Similarly, Table 7 shows the findings of NS GAN on breast cancer dataset. When evaluating the performance of these models on the breast cancer dataset, K-CGAN outperformed NS GAN in all metrics considered. K-CGAN achieved XGBoost's accuracy of 99.125%, which indicates that the model was able to correctly classify 99.125% of the cases in the dataset. In contrast, NS GAN had an accuracy of 89.99.03% on XGBoost, which is considerably lower than K-CGAN. Additionally, K-CGAN achieved high precision, recall, and F1 score values, indicating that it was better at correctly classifying both positive and negative cases in the dataset. Furthermore, when looking at the performance of each individual model used in the analysis, it is clear that K-CGAN performed consistently well across all models, achieving a precision score of at least 0.977995 for all models. On the other hand, NS GAN had varied performances across different models. For instance, MLP achieved the highest F1 score of 0.918033 among all NS GAN models, while MLP achieved the highest accuracy score of 0.988489. The K-CGAN and NS GAN are useful machine learning models for breast cancer classification, but K-CGAN outperformed NS GAN in terms of classification accuracy and consistency across different models. These results demonstrate the potential of machine learning models based on GANs for improving the accuracy and reliability of breast cancer diagnosis.

Table 8 shows the performance of SDG GAN when applied on the same dataset. When comparing the accuracy scores of the two models, K-CGAN shows better performance. The accuracy score of K-CGAN ranges from 0.98875 to 0.99250 for all models, while the accuracy score of SDG GAN ranges from 0.724077 to 0.910369. This indicates that K-CGAN is a more accurate and reliable model for breast cancer classification. Secondly, when comparing the F1 Score of both models, K-CGAN has a higher F1 score range between 0.988875 to 0.992556, while SDG GAN F1 score ranges from 0.645367 to 0.868841. Moreover, the K-CGAN also outperformed SDG GAN in terms of better precision and recall values. K-CGAN indicates a more precise prediction of positive cancer diagnosis as well as better handling of false positives.

Overall, based on the classification performance of each GAN model, K-CGAN demonstrates superior classification performance compared to other GANs applied to breast cancer data. While other GANs illustrated promising results in some models, no other technique showed the same level of consistent success as K-CGAN. Therefore, K-CGAN can be a recommended technique for accurate and reliable breast cancer classification.

From Table 9, it is evident that the performance of K-CGAN outweighs that of the other models by a much higher F1 score.

TABLE 9. Multiple models F1 Score results comparison.

Model	LS GAN	WS GAN	NS GAN	SDG GAN	K-CGAN
XGBoost	0.857741	0.896074	0.795580	0.768362	**0.991326**
Random Forest	0.902326	0.925301	0.927273	0.728843	**0.988875**
Nearest Neighbor	0.901961	0.917073	0.933025	0.866841	**0.991326**
MLP	0.917706	0.952153	0.918033	0.645367	**0.992556**
Logistic Regression	0.907731	0.924939	0.922727	0.863158	**0.988875**

Table 9 show a brief comparison of GAN frameworks. The F1 scores of K-CGAN are higher than those of LS GAN, WS GAN, NS GAN, and SDG GAN. The K-CGAN model yielded high-quality results in terms of its F1 score across all the classifiers. MLP achieved an F1 score of 0.991326, while Random Forest and Logistic Regression both scored 0.988875. Nearest neighbor and XGBoost performed equally well, with F1 scores of 0.991326 and 0.991326, respectively. These results demonstrate the effectiveness of K-CGAN in achieving high scores across a range of classifiers.

In addition, Table 10 compares the precision values of each method.

TABLE 10. Multiple models precision results comparison.

Model	LS GAN	WS GAN	NS GAN	SDG GAN	K-CGAN
XGBoost	0.770677	0.877828	0.960000	0.957747	0.982801
Random Forest	0.889908	0.945813	0.894737	0.574932	0.977995
Nearest Neighbor	**0.938776**	0.949495	0.914027	0.970780	0.982801
MLP	**0.973545**	0.966019	0.911628	1.000000	0.985222
Logistic Regression	**0.962963**	0.950219	0.890351	0.976190	0.977995

As seen from the Table 10, the precision values for the K-CGAN method are improved when compared with other GAN methods. The classification performance was checked on popular classification methods. The precision scores for the K-CGAN method were higher than the precision values of LSGAN, WDGAN, NSGAN and SDGGAN. The K-CGAN method attained higher values in terms of its precision values for all the classifiers. MLP classifier performed the best for K-CGAN. The MLP classifier was followed by XGBoost and Nearest Neighbor as both had a same precision value of 0.982801. The third best precision values were achieved by Random Forest and Logistic Regression at precision value of 0.977995.

In this study, we suggested a K-CGAN trained to produce synthetic data to address insufficient dataset problems in the medical domain.This study used numerical data from 569 patients, of which 212 have malignant tumors and 357 have benign tumors.

VII. CONCLUSION

Furthermore, it was aimed to develop a model that could yield exceptional results in the breast cancer domain. Using the Generator model, we created synthetic data and then compared the data with the original breast cancer dataset. Our empirical findings show that the proposed K-CGAN method performed better than other GANs in terms of classification performance, and the Non-Saturating GAN also achieved good results. We evaluated the classification performance of our proposed GAN using recall, precision, accuracy, and F1 scores on synthetically generated data.

The empirical investigation demonstrated that, when compared to other GAN versions, the K-CGAN performed well and had the highest stability.

REFERENCES

[1] Y.-C. Teh, G.-H. Tan, N. A. Taib, K. Rahmat, C. J. Westerhout, F. Fadzli, M.-H. See, S. Jamaris, and C.-H. Yip, "Opportunistic mammography screening provides effective detection rates in a limited resource healthcare system," *BMC Cancer*, vol. 15, no. 1, p. 405, Dec. 2015, doi: 10.1186/s12885-015-1419-2.

[2] S. Shams, R. Platania, J. Zhang, J. Kim, K. Lee, and S.-J. Park, "Deep generative breast cancer screening and diagnosis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2018, pp. 859–867, doi: 10.1007/978-3-030-00934-2_95.

[3] S. Destounis and A. Santacroce, "Age to begin and intervals for breast cancer screening: Balancing benefits and harms," *Amer. J. Roentgenol.*, vol. 210, no. 2, pp. 279–284, Feb. 2018, doi: 10.2214/AJR.17.18730.

[4] R. Rawal, "Breast cancer prediction using machine learning," *J. Emerg. Technol. Innov. Res.*, vol. 7, no. 5, pp. 13–24, May 2020.

[5] M. Kowal, P. Filipczuk, A. Obuchowicz, J. Korbicz, and R. Monczak, "Computer-aided diagnosis of breast cancer based on fine needle biopsy microscopic images," *Comput. Biol. Med.*, vol. 43, no. 10, pp. 1563–1572, Oct. 2013, doi: 10.1016/j.compbiomed.2013.08.003.

[6] M. Broeders, S. Moss, L. Nyström, S. Njor, H. Jonsson, E. Paap, N. Massat, S. Duffy, E. Lynge, and E. Paci, "The impact of mammographic screening on breast cancer mortality in Europe: A review of observational studies," *J. Med. Screening*, vol. 19, no. 1, pp. 14–25, Sep. 2012, doi: 10.1258/jms.2012.012078.

[7] S. D. Desai, S. Giraddi, N. Verma, P. Gupta, and S. Ramya, "Breast cancer detection using GAN for limited labeled dataset," in *Proc. 12th Int. Conf. Comput. Intell. Commun. Netw. (CICN)*, Sep. 2020, pp. 34–39, doi: 10.1109/CICN49253.2020.9242551.

[8] M.-S. Ong and K. D. Mandl, "National expenditure for false-positive mammograms and breast cancer overdiagnoses estimated at $4 billion a year," *Health Affairs*, vol. 34, no. 4, pp. 576–583, Apr. 2015, doi: 10.1377/hlthaff.2014.1087.

[9] A. D. Trister, D. S. M. Buist, and C. I. Lee, "Will machine learning tip the balance in breast cancer screening?" *JAMA Oncol.*, vol. 3, no. 11, p. 1463, Nov. 2017, doi: 10.1001/jamaoncol.2017.0473.

[10] S. Ara, A. Das, and A. Dey, "Malignant and benign breast cancer classification using machine learning algorithms," in *Proc. Int. Conf. Artif. Intell. (ICAI)*, Apr. 2021, pp. 97–101, doi: 10.1109/ICAI52203.2021.9445249.

[11] W. Lotter, G. Sorensen, and D. Cox, "A multi-scale CNN and curriculum learning strategy for mammogram classification," in *Proc. Int. Workshop Deep Learn. Med. Image Anal.*, Jul. 2017, pp. 169–177.

[12] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[13] M. Elhoseiny, Y. Zhu, H. Zhang, and A. Elgammal, "Link the head to the 'beak': Zero shot learning from noisy text description at part precision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6288–6297.

[14] J. Xie, R. Liu, J. Luttrell, and C. Zhang, "Deep learning based analysis of histopathological images of breast cancer," *Frontiers Genet.*, vol. 10, pp. 1–12, Feb. 2019, doi: 10.3389/fgene.2019.00080.

[15] C. Guttà, C. Morhard, and M. Rehm, "Applying GAN-based data augmentation to improve transcriptome-based prognostication in breast cancer," *medRxiv*, vol. 2022, pp. 1–29, Jan. 2022.

[16] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," 2017, *arXiv:1702.05747*.

[17] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020, doi: 10.1145/3422622.

[18] E. Wu, K. Wu, D. Cox, and W. Lotter, "Conditional infilling GANs for data augmentation in mammogram classification," in *Proc. Int. Workshop Reconstruction Anal. Moving Body Organs*, Jul. 2018, pp. 1–8.

[19] J. M. Wolterink, A. M. Dinkla, M. H. F. Savenije, P. R. Seevinck, C. A. T. van den Berg, and I. Išgum, "Deep MR to CT synthesis using unpaired data," in *Proc. Int. Workshop Simulation Synth. Med. Imag.*, 2017, pp. 14–23, doi: 10.1007/978-3-319-68127-6_2.

[20] D. Nieet, "Medical image synthesis with context-aware generative adversarial networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 10435, Sep. 2017, pp. 417–425, doi: 10.1007/978-3-319-66179-7_48.

[21] J. T. Guibas, T. S. Virdi, and P. S. Li, "Synthetic medical images from dual generative adversarial networks," 2017, *arXiv:1709.01872*.

[22] H. Salehinejad, S. Valaee, T. Dowdell, E. Colak, and J. Barfett, "Generalization of deep neural networks for chest pathology classification in X-rays using generative adversarial networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 990–994, doi: 10.1109/ICASSP.2018.8461430.

[23] L. Hou, A. Agarwal, D. Samaras, T. M. Kurc, R. R. Gupta, and J. H. Saltz, "Unsupervised histopathology image synthesis," 2017, *arXiv:1712.05021*.

[24] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "Synthetic data augmentation using GAN for improved liver lesion classification," in *Proc. 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 289–293.

[25] D. Ribli, A. Horváth, Z. Unger, P. Pollner, and I. Csabai, "Detecting and classifying lesions in mammograms with deep learning," *Sci. Rep.*, vol. 8, no. 1, p. 4165, Mar. 2018, doi: 10.1038/s41598-018-22437-z.

[26] S. Shao, P. Wang, and R. Yan, "Generative adversarial networks for data augmentation in machine fault diagnosis," *Comput. Ind.*, vol. 106, pp. 85–93, Apr. 2019, doi: 10.1016/j.compind.2019.01.001.

[27] S. Pandey, P. R. Singh, and J. Tian, "An image augmentation approach using two-stage generative adversarial network for nuclei image segmentation," *Biomed. Signal Process. Control*, vol. 57, Mar. 2020, Art. no. 101782, doi: 10.1016/j.bspc.2019.101782.

[28] N. K. Singh and K. Raza, "Medical image generation using generative adversarial networks: A review," in *Health Informatics: A Computational Perspective in Healthcare*, 2021, pp. 77–96, doi: 10.1007/978-981-15-9735-0_5.

[29] M. Alauthman, A. Al-Qerem, B. Sowan, A. Alsarhan, M. Eshtay, A. Aldweesh, and N. Aslam, "Enhancing small medical dataset classification performance using GAN," *Informatics*, vol. 10, no. 1, p. 28, Mar. 2023, doi: 10.3390/informatics10010028.

[30] T. Suresh, Z. Brijet, and T. D. Subha, "Imbalanced medical disease dataset classification using enhanced generative adversarial network," *Comput. Methods Biomechanics Biomed. Eng.*, vol. 2022, pp. 1–17, Nov. 2022, doi: 10.1080/10255842.2022.2134729.

[31] M. Abedi, L. Hempel, S. Sadeghi, and T. Kirsten, "GAN-based approaches for generating structured data in the medical domain," *Appl. Sci.*, vol. 12, no. 14, p. 7075, Jul. 2022, doi: 10.3390/app12147075.

[32] R. Sauber-Cole and T. M. Khoshgoftaar, "The use of generative adversarial networks to alleviate class imbalance in tabular data: A survey," *J. Big Data*, vol. 9, no. 1, p. 98, Aug. 2022, doi: 10.1186/s40537-022-00648-6.

[33] M. S. K. Inan, S. Hossain, and M. N. Uddin, "Data augmentation guided breast cancer diagnosis and prognosis using an integrated deep-generative framework based on breast tumor's morphological information," *Informat. Med. Unlocked*, vol. 37, Jan. 2023, Art. no. 101171, doi: 10.1016/j.imu.2023.101171.

[34] A. Mert, "Enhanced dataset synthesis using conditional generative adversarial networks," *Biomed. Eng. Lett.*, vol. 13, no. 1, pp. 41–48, Feb. 2023, doi: 10.1007/s13534-022-00251-x.

[35] E. Wu, K. Wu, and W. Lotter, "Synthesizing lesions using contextual GANs improves breast cancer classification on mammograms," 2020, *arXiv:2006.00086*.

[36] M. Saini and S. Susan, "Deep transfer with minority data augmentation for imbalanced breast cancer dataset," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 106759, doi: 10.1016/j.asoc.2020.106759.

[37] D. Bardou, K. Zhang, and S. M. Ahmad, "Classification of breast cancer based on histology images using convolutional neural networks," *IEEE Access*, vol. 6, pp. 24680–24693, 2018, doi: 10.1109/ACCESS.2018.2831280.

[38] S. Zhou, U. J. Islam, N. Pfeiffer, I. Banerjee, B. K. Patel, and A. Iquebal, "SCGAN: Sparse CounterGAN for counterfactual explanations in breast cancer prediction," *medRxiv*, vol. 2023, pp. 1–4, Jun. 2023.

[39] E. Akkur, F. Türk, and O. Erogul, "Breast cancer classification using a novel hybrid feature selection approach," *Neural Netw. World*, vol. 33, no. 2, pp. 67–83, 2023.

[40] K. T. Chui, B. B. Gupta, R. H. Jhaveri, H. R. Chi, V. Arya, A. Almomani, and A. Nauman, "Multiround transfer learning and modified generative adversarial network for lung cancer detection," *Int. J. Intell. Syst.*, vol. 2023, pp. 1–14, Mar. 2023, doi: 10.1155/2023/6376275.

[41] F. Zhang, Y. Zhang, X. Zhu, X. Chen, H. Du, and X. Zhang, "PregGAN: A prognosis prediction model for breast cancer based on conditional generative adversarial networks," *Comput. Methods Programs Biomed.*, vol. 224, Sep. 2022, Art. no. 107026, doi: 10.1016/j.cmpb.2022.107026.

[42] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*.

[43] M. Shannon, B. Poole, S. Mariooryad, T. Bagby, E. Battenberg, D. Kao, D. Stanton, and R. Skerry-Ryan, "Non-saturating GAN training as divergence minimization," 2020, *arXiv:2010.08029*.

[44] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2813–2821, doi: 10.1109/ICCV.2017.304.

[45] C. Charitou, S. Dragicevic, and A. d'Avila Garcez, "Synthetic data generation for fraud detection using GANs," 2021, *arXiv:2109.12546*.

[46] W. H. Wolberg, W. N. Street, and O. L. Mangasarian. (1992). *Breast Cancer Wisconsin (Diagnostic) Data Set*. UCI Machine Learning Repository. [Online]. Available: http://archive.ics.uci.edu/ml/

[47] T. M. Mitchell, *Machine Learning*, vol. 1. New York, NY, USA: McGraw-Hill, 1997.

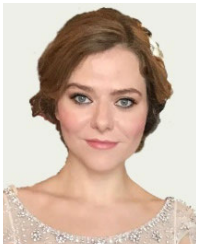
[48] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[49] D. G. Kleinbaum and M. Klein, *Logistic Regression*. New York, NY, USA: Springer, 2010, doi: 10.1007/978-1-4419-1742-3.

[50] T. Chen and T. He, "Xgboost: Extreme gradient boosting," Tech. Rep., 2022.

[51] L. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009, doi: 10.4249/scholarpedia.1883.

[52] H. Taud and J. F. Mas, "Multilayer perceptron (MLP)," in *Geomatic Approaches for Modeling Land Change Scenarios*, 2018, pp. 451–455, doi: 10.1007/978-3-319-60801-3_27.



EMILIJA STRELCENIA is currently a Ph.D. Researcher with Bournemouth University. She has presented her findings to a variety of audiences, including conferences and corporate meetings and publishing in peer-reviewed journals. Her current research interests include AI, cyber security, and internet crime prevention. She has earned several awards in acknowledgement of her invaluable contributions. Her accomplishments have been recognized by numerous international conferences, including IEEE Third International Conference on Artificial Intelligence Technology (CAIT 2022), in 2022, IEEE 12th International Conference on Power and Energy Systems (ICPES 2022), in 2022, and IEEE Third International Conference on Computing, Networking, Telecommunications and Engineering Sciences Applications (CoNTESA 2022), in 2022.

SIMANT PRAKOONWIT received the B.Eng. degree from Chulalongkorn University, Bangkok, and the M.Sc. and Ph.D. degrees from Imperial College London. He was a Postdoctoral Research Assistant with Imperial College London, where he focuses on 3D computer vision projects funded by the Home Office, U.K., with applications in security and medical fields. He has held academic positions at several U.K. universities. He is currently an Associate Professor with Bournemouth University, where he is teaching AI and working on funded AI-related research projects. He has made significant contributions to the fields of AI and 3D computer vision through his teaching and consulting work in academia and government organizations. His current research interests include AI and 3D computer vision. He has won a British Government Scholarship to study in the U.K. He has won numerous research awards, including the IEEE Innovation and Creativity Award, the Brunel Research Initiative Enterprise Fund Award, and the best papers at conferences and journal special issues.