Do there exist an emotion trend in scientific papers? PRO-VE conference as a case

Rishitha Venumuddala, Lai Xu and Paul de Vrieze

Department of Computing and Informatics, Bournemouth University, Poole, BH12 5BB Bournemouth, United Kingdom

{s5536852,lxu,pdvrieze}@bournemouth.ac.uk

Abstract. Scientific writing aims for formality and objectivity, yet emotions are integral to human communication, decision-making, and collaboration, all of which are fundamental to scientific progress. Existing research on emotion detection has mainly focused on datasets from social media and online platforms, where emotional expressions are abundant. However, scientific texts pose unique challenges due to their formal language and the rarity of explicit emotional words, necessitating specialised investigation. This study investigates the presence and nature of emotions in scientific texts, specifically analysing the abstracts from the PRO-VE conference series from 2012 to 2022. Two emotion detection methods are employed: a lexicon-based approach and a hybrid machine learning-based approach. The lexicon-based approach utilises the NRC Emotion Lexicon to identify and quantify emotions within the PRO-VE abstracts, while the hybrid approach integrates Word2Vec for word embedding generation and a Random Forest classifier for emotion prediction. The findings reveal a predominance of positive emotions, such as trust, anticipation, and joy, in the PRO-VE abstracts, consistent with the objective nature of scientific writing. In light of the PRO-VE conference series' 25th anniversary, an analysis of trends and patterns in the detected emotions offers insights into the emotional landscape of this prestigious conference series. The study also critically examines the limitations of the experiments, including the dataset size and the prevalence of positive emotions.

Keywords: Emotion Detection· PRO-VE· Emotion in Collaborative Networks· Emotion Trend in Collaborative Networks, Collaborative Networks

1 Introduction

Various forms of Collaborative Networks (CNs), such as virtual enterprises (VEs), virtual organisations (VOs), and dynamic supply chains, have emerged in the last few decades to address ever-evolving challenges in scientific and business fields. In CNs, participating entities (organisations and individuals) are heterogeneous in terms of culture, needs, goals, and geographical distribution, yet they collaborate to achieve common objectives. The PRO-VE working conference is a collaborative network of virtual enterprises that aims to address business needs by encouraging collaboration and knowledge sharing among researchers from multiple fields, including computing, social science, manufacturing, and others [1].

CNs are inherently complex due to the involvement of diverse participants with varying thoughts and expectations, collaborating to achieve shared goals. In this process, there is a possibility of generating emotions that could affect the dynamics of the CN in achieving its purpose [2].

Research on emotions reveals their potential as drivers of decision-making. Emotions influence individuals' choices, judgments, and decisions, playing a vital role in communication, network formation, collaboration, and knowledge sharing within groups [3]. The emotions of every participating entity contribute to the overall emotion conveyed by a Collaborative Network.

Since 1999, the PRO-VE conference series has encouraged researchers worldwide to contribute to collaborative networks, addressing evolving challenges in computing, social sciences, manufacturing, and other fields. Over the past two decades, more than 1,000 conference papers have been published in these series by over 200 researchers. This implies that the emotions of every single author and the emotions conveyed in every abstract contribute to the overall emotion conveyed by the PRO-VE conference.

While many existing emotion datasets are derived from social media (e.g., tweet datasets from Twitter and other online platforms) and are rich in emotional words originating from user utterances, scientific papers have formal text without emotion-rich words, making them distinct from existing emotion datasets.

An exploratory research methodology is followed due to limited existing research on emotion detection in scientific papers compared to other prominent fields such as human-computer interaction (HCI), social media analysis, services industry, and healthcare. This research aims to discover whether emotions exist in scientific research papers, typically written in a neutral manner, by identifying the emotions conveyed by the authors of PRO-VE conference papers to readers through their abstracts.

This research endeavours to gain insights into the emotional impact of scientific papers and contribute to the growing research on emotional disclosure in the scientific arena. The research questions are:

1. Are emotions present in scientific papers from PRO-VE conferences? If present, what are the dominant emotions being conveyed by the collaborative network of the PRO-VE conference through their abstracts?
2. Are there any trends in the identified emotions of the PRO-VE abstract dataset?

By revealing the emotional aspects of scientific writing within the context of the PRO-VE conference, this research enhances the understanding of emotional expression in academic research, particularly in computer science and related fields. The findings offer valuable insights into the emotional dimensions of scientific writing, highlighting how emotions influence scientific communication, collaboration, and knowledge dissemination. This study paves the way for deeper exploration of the role emotions play in academic discourse.

The structure of this paper is organised as follows: Section 2 reviews existing research related to emotion detection and associated methods; Section 3 introduces the design details of the proposed methods employed in this study; Section 4 outlines the experimental setup, including the datasets, preprocessing steps, and evaluation metrics used; Section 5 presents a comprehensive discussion of the experimental results, analysing the performance of the proposed methods; and finally, Section 6 concludes

the study by summarising the key findings, highlighting the contributions, and suggesting potential directions for future research.

2 Related work

Emotions play a vital role in human behaviour and decision-making. Research on emotions has come a long way, but there is still no consensus on the definition of emotion in literature [4] or the categorisation of emotions [5]. Understanding and detecting or recognising emotions in text, speech, facial expressions, etc., is crucial for human research [5] and is a complex process. Numerous papers have been published by researchers on emotion detection from text in various fields, including computer science, psychology, and medicine.

Some prominent theories on identifying and classifying emotions are Ekman's theory of six emotions and Plutchik's eight-wheel emotion theory. Ekman argued that there are six basic emotions: anger, fear, surprise, disgust, happiness, and sadness[6]. On the other hand, Plutchik proposed the "wheel of eight emotions," which includes all six emotions from Ekman's theory, as well as anticipation and trust [7]. The emotions are arranged in such a way that opposing emotions are placed opposite to each other: joy versus sadness, trust versus disgust, fear versus anger, and surprise versus anticipation.

Emotions are expressed through text, speech, and facial expressions. Text is considered the primary choice for emotion expression in everyday life, at work, or in social events [8]. Determining the positive, negative, and neutral polarity from text is called sentiment analysis, and analysing emotions from text is known as emotion analysis [9]. Extracting emotions from text is a difficult task, as one word can have multiple meanings and different emotions, and multiple words can have similar emotions. Research on extracting emotions from text has increased tremendously, and multiple researchers have contributed to this field due to the increased usage of social media and other online platforms like TripAdvisor and Amazon. Numerous papers on emotion detection have been published on mental health analysis in social media [8], the effect of customer reviews and social media posts on sales and market analysis, brand profiling, etc. [10].

Gievska *et.al.* proposed a hybrid model based on lexical analysis and machine learning for classifying Ekman's six emotions in user's text for affective interaction [11]. They performed experiments to test the hybrid approach's performance against individual approaches (NRC lexicon and SVM classifier). The hybrid approach outperformed the individual approaches in terms of performance and reduction of text misclassification. It obtained a precision score of 85.4, recall of 84.6, and F-score of 84.6, whereas the machine learning approach obtained precision of 54.5, recall of 54.0, and F-measure of 52.8, and lexical analysis achieved precision of 74.6, recall of 72.1, and F-measure of 73.7. However, this research used only shorter conversation texts.

Sarsam *et. al.* proposed a lexicon-based approach to detect suicide-related data from tweets on social media, specifically Twitter[12]. They used the NRC Affect Lexicon and SentiStrength to identify users' sentiments and emotions. Subsequently, a semi-supervised approach was designed by incorporating the YATSI classifier to predict suicide-related tweets. The emotions obtained from the mentioned lexicons enhanced the predictive capability of the classifier. Fear, anger, and negative sentiments were

found in suicide-related tweets. However, a limitation of this approach is that only emotions were calculated, and adding sentiment-related features could enhance the classification.

Park *et.al.* proposed an Emotion Embedding Model (EEM) for emotion detection from text stories using a supervised learning algorithm, CNN[13]. The proposed model utilised tweets with emotional hashtags as labels and the ROC Stories dataset. Emotions were classified into Plutchik's eight emotion categories using CNN. The emotion "joy" obtained the highest accuracy of 73.3%, while "anger" resulted in the lowest accuracy of 36.7%. A limitation of this approach is that contextual information spanning multiple sentences and the negation of sentences (e.g., "not," "never," "little") were not taken into account.

Pouransari and Ghili proposed a sentiment analysis approach for movie reviews using Word2Vec and supervised machine learning classifiers, namely SVM, Random Forest, Logistic Regression, and Recurrent Neural Networks (RNTN) [14]. The IMDB movie review dataset was used in this approach. The model built using Word2Vec and the Random Forest classifier attained an accuracy of 84%, outperforming other classifiers. In the second approach, a low-rank RNTN performed better than a high-rank RNTN with the same accuracy but less computational time.

Deho *et. al.*performed sentiment analysis on tweets related to a US military base in [15]. A hybrid model was built using Word2Vec and the Random Forest classifier to perform sentiment analysis. The proposed model achieved an accuracy of 81% and performed well in detecting tweets with positive sentiments due to the good quality of word vectors.

3 Analysis methods

In this section, we introduce the data collection process in Section 3.1 and discuss the detailed design of the proposed methods for emotion detection in Section 3.2.

3.1 Data preparation

This research utilises two datasets of abstracts from the PRO-VE Conference. The first dataset encompasses abstracts from 2018 to 2022, while the second includes a broader range from 2012 to 2022. Both datasets were conveniently downloaded from the Scopus database[1] in CSV format, eliminating the need for additional formatting. Each dataset originally contained four columns: link to the paper in Scopus, the abstract itself, author keywords, and indexed keywords. However, our analysis focuses solely on the abstracts, so the author and indexed keyword columns were excluded.

To ensure data quality, each dataset underwent verification to identify and address any anomalies. These anomalies could include extra columns accidentally downloaded from conferences outside the intended year range, or missing abstracts from downloaded conference proceedings that typically lack them. In these cases, the dataset preparation step was repeated to remove irrelevant data and obtain a clean dataset for further analysis. It's important to note that minimal preprocessing was required due to

[1] https://www.scopus.com/sources.uri?zone=TopNavBar&origin=

the nature of the data. The Scopus-provided CSV format and the relatively simple text within the abstracts themselves reduced the need for tasks like noise removal or normalisation.

3.2 Workflow

First, we employed the Lexicon approach with LeXmo (NRC Emotion Lexicon) to analyse emotions in PRO-VE conference abstracts (2018-2022). LeXmo classifies words into eight emotions and two sentiment valences (positive/negative) using its 14,000-word and 25,000-word sense lexicon. Within a Jupyter notebook environment, LeXmo analysed each abstract to calculate emotion scores. We then determined the range of emotions (lowest and highest scores) within each abstract and visualised average emotion scores across all abstracts using line graphs. This process was repeated for each year's dataset to gain insights into emotional trends over time. Figure 1 depicts the workflow of our experiments.

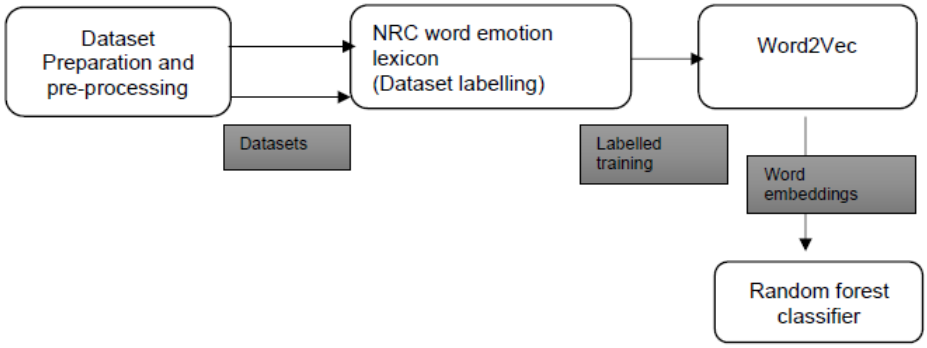

Figure 1 Workflow of our experiments.

Following the calculation of emotion scores for each abstract in the dataset using the NRC word emotion lexicon, a new field named 'dominant emotion' was added. This field identifies the emotion with the highest score within each abstract. Essentially, each record was annotated with its most prominent emotion based on the emotion scores. Finally, a single training dataset was created by concatenating all datasets from 2018 to 2021.

Traditional machine learning models, such as linear regression, logistic regression, decision trees, and neural networks, excel at predicting based on word frequency and probability. However, they often struggle to capture the context in which words appear. To overcome this limitation, we propose a hybrid approach that leverages Word2Vec and a Random Forest classifier.

Word2Vec generates word embeddings, which represent the semantic relationships between words. This allows the model to go beyond simple word counts and consider the meaning within the text. We applied Word2Vec to our labelled training dataset. Each abstract was converted into a word embedding. These embeddings were then used to train and test the Random Forest classifier model.

4 Experiments

Experiments are conducted on two different datasets. For each dataset, two emotion detection methods are employed. Section 4.1 presents the experiments using the 2018-2022 dataset, while Section 4.2 covers the experiments using the 2012-2022 dataset.

4.1 Experiment using 2018-2022 dataset

Lexicon approach

After calculating emotion scores for each abstract in the 2018-2022 dataset, the average scores were calculated for the entire dataset to determine the highest and lowest emotions. Disgust emerged as the least prevalent emotion, followed by anger, while trust was the highest, closely followed by anticipation. Notably, positive valence was significantly higher than negative valence for the year 2022.

Table 1 displays the average emotion values for all the years from 2018 to 2022 using lexicon approach. Trust was the highest emotion across all years, with its values remaining relatively stable. In 2018, the trust value was 0.0427, which increased to 0.0464 in 2019 but slightly decreased to 0.0356 in 2020. It then rose to 0.0398 in 2021 and finally reached 0.0419 in 2022. Conversely, disgust consistently remained the least prevalent emotion, with values less than or equal to 0.0005, except for 2020 when it was 0.0008, though the difference was negligible.

Table 1 Average emotion scores from 2018 to 2022 based on Lexicon approach

Emotions and valence	2018	2019	2020	2021	2022
Anger	0.0021	0.0018	0.0022	0.0017	0.0020
Anticipation	0.0279	0.0316	0.0261	0.0253	0.0283
Disgust	0.0002	0.0005	0.0008	0.0005	0.0005
Fear	0.0063	0.0068	0.0075	0.0059	0.0075
Joy	0.0109	0.0131	0.0102	0.0099	0.0101
Negative	0.0100	0.0113	0.0126	0.0094	0.0118
Positive	0.0406	0.0452	0.0371	0.0405	0.0425
Sadness	0.0041	0.0039	0.0051	0.0040	0.0051
Surprise	0.0039	0.0049	0.0038	0.0065	0.0049
Trust	0.0427	0.0464	0.0356	0.0398	0.0419

There were some fluctuations in the values of anger, which stood at 0.0020 in 2018, decreased in 2019, increased again in 2020, decreased in 2021, and finally rose to 0.0020 in 2022, almost equal to the 2018 value. Similarly, anticipation was the second-highest emotion after trust, remaining almost equal to and below 0.0300 in all years except 2019.

Fear was highest in 2022 at 0.0075 and lowest in 2021 at 0.0059, while sadness peaked at 0.0051 in 2020 and 2022 and was lowest at 0.0039 in 2019. Regarding valence, positive valence was predominant, reaching its highest value of 0.0425 in 2022

and its lowest at 0.0371 in 2020. Negative valence, on the other hand, was highest in 2020 at 0.0126 and lowest in 2018 at 0.0100.

While trust was the predominant emotion from 2018 to 2022, followed by anticipation, and disgust was the least prevalent emotion, with positive valence being the highest, it is noteworthy that negative valence, although lower, cannot be considered negligible. The value of negative valence was almost equal to 0.01, which is greater than zero, indicating the presence of negative emotions in the abstracts.

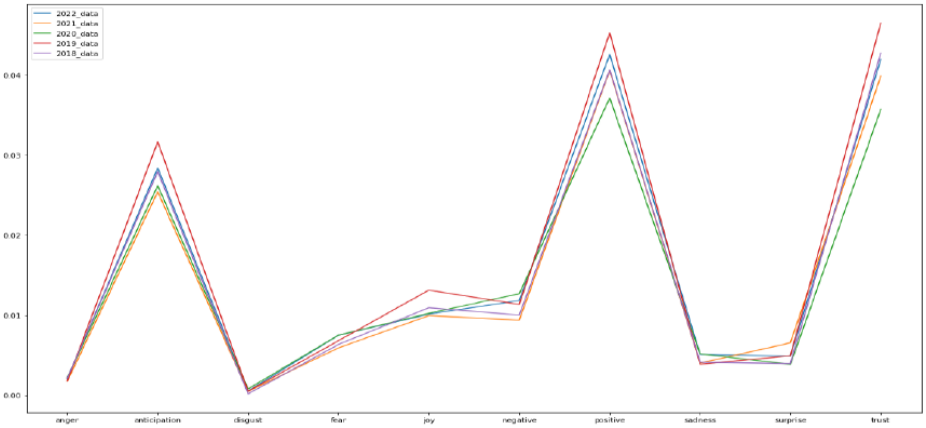

Figure 2 Overall emotions from 2018 to 2022

The line graph plotted against all emotion values for the five years clearly in Figure 2 shows that 2019 had the highest values for anticipation, trust, joy, and positive valence. Disgust and anger, on the other hand, were almost indistinguishable, with negligible differences across all years. Positive valence was lowest in 2020, while negative valence was highest during that year.

Hybrid machine learning approach

After employing the Lexicon approach to calculate emotion scores for each abstract in the 2018-2022 dataset, the dataset labelled using the NRC emotion lexicon was used as input to train and test a proposed machine learning model. This model was built using Word2Vec for creating word embeddings and a Random Forest classifier to predict emotions from these word embeddings.

The dataset consisted of 217 records, with most of them labelled as trust, anticipation, and positive, while only a few were labelled as fear. Consequently, the model was trained to predict only these four emotions in the validation dataset. Emotions such as surprise, disgust, sadness, and joy were not present in the dataset.

The hybrid machine learning approach achieved an accuracy of 68%. Additionally, the performance metrics of precision, recall, and F1-score were calculated and presented in Table 2.

Table 2 Performance of hybrid model of 2018-2022 dataset

	Precision	**Recall**	**F1-score**

Negative	100%	100%	100%
Anticipation	100%	59%	70%
Trust	71%	67%	69%
Positive	55%	86%	67%

All performance metrics for the Negative emotion were 100%, indicating that the model accurately predicted all abstracts with negative valence as the dominant emotion without any mispredictions. Anticipation attained 100% precision, 59% recall, and a 70% F1-score, suggesting that the model was highly precise in identifying anticipation but may have missed some instances. Trust achieved a precision of 71%, a recall of 67%, and an F1-score of 69%, indicating a balanced performance in predicting this emotion.

Positive valence obtained a precision of 71%, a recall of 67%, and an F1-score of 69%, similar to the performance for trust, while negative valence obtained the highest performance, with all abstracts containing negative valence as the dominant emotion being accurately predicted without any mispredictions.

4.2 Experiment using 2012-2022 Dataset

Initially, the training dataset from 2018 to 2021 contained only 218 records, which was a relatively small size. To measure the performance of the model more accurately, the size of the training dataset was increased by incorporating data from 2012.

The same procedure was followed to create the datasets from 2012. The new training dataset, spanning from 2012 to 2022, was given as input to the NRC Emotion lexicon for annotating each record with the dominant emotion. Emotion scores were calculated for each abstract in the dataset, and a new attribute, "dominant emotion," was added and annotated with the highest emotion score for that record. As a result, the expanded 2012-2022 dataset contained 718 annotated records, which could be used to train and test the hybrid machine learning model.

The training dataset was given as input to Word2Vec, which generated 718-word embeddings. These word embeddings were then provided as input to a Random Forest classifier to train and test the model. Most records were labelled with trust, anticipation, and positive, as they were the dominant emotions, while only a few were labelled with negative and fear. It's worth noting that the previous dataset (2018-2021) had zero records with negative valence, but by increasing the size of the dataset, negative valence was added to the training dataset.

Despite the larger dataset, the model obtained an accuracy of 60%, which was 8% less compared to the smaller dataset. The performance metrics of emotions are mentioned in Table 3.

Table 3 Performance of hybrid model of 2012-2022 dataset

	Precision	**Recall**	**F1-score**
Anticipation	100%	22%	36%
Negative	100%	50%	67%
Trust	81%	38%	52%

Positive	54%	97%	69%

Anticipation and negative obtained 100% precision, followed by trust at 81%. Anticipation achieved 100% precision, 22% recall, and a 36% F1-score, implying that the model was highly accurate in recognising anticipation but overlooked numerous instances. Negative secured 100% precision, 50% recall, and a 67% F1-score, exhibiting a balanced performance in predicting negative emotions. Trust attained a precision of 81%, a recall of 38%, and an F1-score of 52%, indicating a relatively low ability to identify all relevant instances but high accuracy in the instances it did identify. Positive achieved a precision of 54%, a recall of 97%, and an F1-score of 69%, suggesting that the model was highly successful in detecting positive emotions but had a lower degree of accuracy.

Balanced the class weights for hybrid machine learning approach

Multiple experiments were performed to improve the performance of the model with the large dataset. As mentioned in the previous experiments, the dataset was imbalanced, with emotion classes not equally distributed. Most records were annotated as positive, followed by trust and anticipation, while very few were labelled as negative and fear. This created a significant imbalance in the dataset.

To address this issue, the class weight parameter of the Random Forest classifier was set to "balanced" during training. The model trained with the balanced dataset attained an accuracy of 59%, which was one percent less than the accuracy of the model trained with the imbalanced dataset. The performance metrics of emotions with balanced class weights are showed in Table 4.

Table 4 Performance of hybrid model with balanced dataset of 2012-2022 dataset

	Precision	**Recall**	**F1-score**
Anticipation	100%	28%	44%
Negative	100%	50%	67%
Trust	86%	27%	41%
Positive	53%	98%	68%

Anticipation and negative scored 100% precision, followed by trust at 86%. Anticipation attained 100% precision, 28% recall, and a 44% F1-score, signifying that the model was highly precise in identifying anticipation but failed to capture many instances. Negative obtained 100% precision, 50% recall, and a 67% F1-score, demonstrating a balanced performance in predicting negative emotions. Trust achieved a precision of 86%, a recall of 27%, and an F1-score of 41%, indicating a relatively low capability to identify all relevant instances but high accuracy in the instances it did identify. Positive secured a precision of 53%, a recall of 98%, and an F1-score of 68%, implying that the model was highly successful in detecting positive emotions but had a lower degree of accuracy.

While the experiments were performed to improve the model's performance, the results showed a contrasting trend. The model attained higher accuracy for the small dataset and lower accuracy for the large balanced dataset.

5 Discussion

5.1 Lexicon approach

Experiment with 2018-2022 dataset

The NRC Emotion Lexicon (EmoLex) is used to identify the emotions associated with each word in an abstract by searching for words annotated with Plutchik's eight emotions: anger, disgust, fear, sadness, joy, anticipation, surprise, and trust. The dictionary of EmoLex contains approximately 14,000 words and 25,000-word tenses. Scores for each emotion are calculated for an abstract in a dataset using EmoLex. The average of each emotion score is used to analyse the distribution of emotions in a dataset from 2018 to 2022, as mentioned in Table 1.

From Table 1, it is evident that trust is the dominant emotion across all years. It was lowest in 2020 at 0.356 and highest in 2019 at 0.464, significantly greater than other emotions in the dataset. Anticipation followed, with the highest score in 2019 at 0.0316 and the lowest in 2021 at 0.0253. The third dominant emotion was joy, peaking in 2019 at 0.0131 and lowest in 2021 at 0.0099. Disgust was the least prevalent emotion, with scores less than 0.0005 in all years, followed by anger, which was less than 0.0022. Fear, sadness, and surprise never exceeded 0.0075. Notably, positive sentiment scores were far greater than negative sentiment scores across all years.

Trust is clearly the dominant emotion in the abstracts of PRO-VE, followed by anticipation and joy. Disgust is the least prevalent emotion, followed by anger. The group {Trust, Anticipation, Joy} represents the dominant emotions, while {Disgust, Anger} is the least dominant group. According to Plutchik's wheel of eight emotions, opposing emotions are placed against each other, with Trust versus Disgust. Trust is the highest emotion across all years, while disgust is the lowest. Positive valence is greater than negative valence for all years, and the dominant group of emotions conveyed is also positive, i.e., {trust, anticipation, joy}.

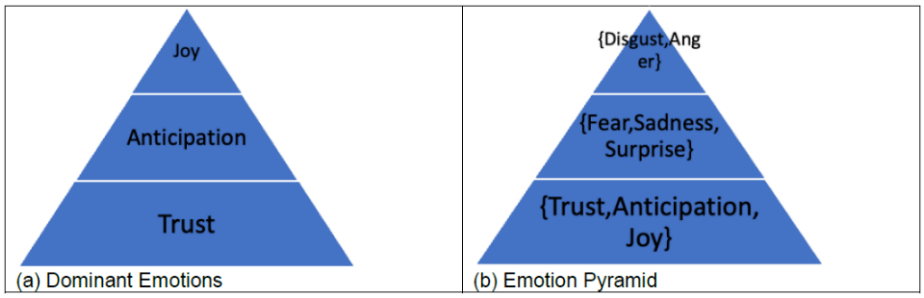

Figure 3 Dominant Emotions and Emotion Pyramid from 2018 to 2022

Figure 3(a) shows that {Trust, Anticipation, Joy} is the set of dominant emotions across all years, followed by {Fear, Sadness, Surprise}, whose values are always close to each other and never greater than 0.0075, whereas {Disgust, Anger} are the least prevalent emotions across all years.

Regarding year-wise analysis, 2019 had the highest {Trust, Anticipation, Joy} scores across all years, whereas 2020 had the lowest trust and positive valence scores and the highest {Disgust, Anger} scores in Figure 3(b). The outbreak of COVID-19 worldwide in 2020 might be the reason for more negative emotions and negative valence compared to other years.

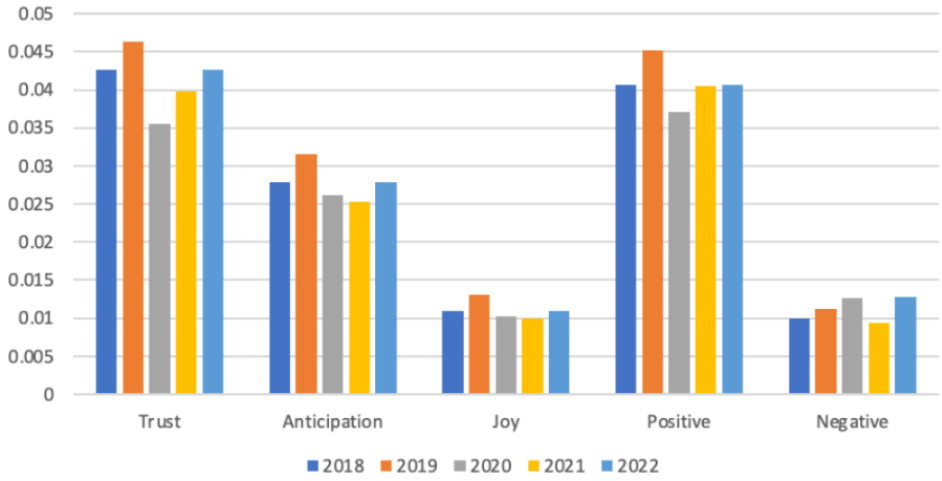

Figure 4 Variation in emotions during COVID from 2018 to 2022

Figure 4 clearly shows that trust, anticipation, joy, and positive valence decreased from 2019 to 2020. 2019 had the highest values for the dominant set {Trust, Anticipation, Joy}, whereas 2020 had the lowest values for the same set across all five years. As mentioned, the reason for the decline in positive valence and the lowest values of the dominant set {Trust, Anticipation, Joy} might be the outbreak of COVID-19 in 2020, which might have had an emotional impact on the authors. Trust and positive valence increased in the following year, 2021, whereas anticipation and joy continued to decline in 2021 and increased in 2022.

On the other hand, negative valence increased from 2019 to 2020 as positive valence decreased. The emotion values of disgust increased from 0.0005 to 0.0008, anger from 0.0018 to 0.0022, fear from 0.0068 to 0.0075, sadness from 0.0034 to 0.0051, and negative valence from 0.0113 to 0.0126 from 2019 to 2020. This means that 2020 had the highest disgust, anger, fear, sadness, and negative valence scores across all five years from 2018 to 2022. The emotion score of surprise also decreased from 0.0049 in 2019 to 0.0038 in 2020, which is the lowest value across all years.

To summarise, the distribution of emotions from 2018 to 2022 shows that 2019 had the highest trust, anticipation, joy, and positive valence scores, whereas 2020 had the lowest values for these emotions. The values of negative emotions like anger, disgust, fear, sadness, and negative valence increased from 2019 to 2020, while the values of

positive emotions like trust, anticipation, joy, surprise, and positive valence scores decreased. This might be due to the outbreak of the Coronavirus and the subsequent mass quarantine and lockdown in 2020. According to studies by Salari et al. (2020), Rajkumar (2020), and Dubey et al. (2020), COVID-19 had a significant impact on public mental health due to various factors, including death, the spread of infection, unavailability of medicine, mass lockdown, and quarantine. Stress, anxiety, and depression increased among people due to the outbreak.

Experiment with 2012-2022 dataset

Table 5 shows that trust is the dominant emotion across all years from 2012 to 2017, followed by anticipation and joy, similar to the years 2018 to 2022 mentioned in Table 1. The highest value of trust is 0.0414 in 2017, while the lowest value is 0.0356 in 2015. Anticipation peaked in 2012 at 0.0312 and was lowest in 2014 at 0.0298. The third most prominent emotion was joy, highest in 2017 at 0.0127 and lowest in 2014 at 0.0091. Disgust remained the least prevalent emotion, consistent with the years 2018 to 2022. It was lowest in 2012 and 2014 at 0.0003, the least emotion score in both datasets. Anger followed as the second-least emotion, peaking in 2014 at 0.0021 and lowest in 2012 at 0.0007. Positive valence in this dataset was also significantly greater than negative valence, with the highest and lowest values being 0.0472 in 2013 and 0.0386 in 2017, respectively. Negative valence was highest in 2016 at 0.0112 and lowest in 2012 at 0.0071.

Table 5 Average emotion scores from 2012 to 2017 using Lexicon approach

Emotions and valences	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**
Anger	0.0007	0.0017	0.0021	0.0014	0.0022	0.0014
Anticipation	0.0312	0.0325	0.0298	0.0281	0.0302	0.0301
Disgust	0.0003	0.0004	0.0003	0.0005	0.0007	0.0004
Fear	0.0057	0.0052	0.0059	0.0069	0.0064	0.0063
Joy	0.0107	0.0117	0.0091	0.0093	0.0103	0.0127
Negative	0.0071	0.0087	0.0098	0.0108	0.0112	0.0093
Positive	0.0440	0.0472	0.0440	0.0431	0.0434	0.0386
Sadness	0.0042	0.0046	0.0040	0.0039	0.0055	0.0031
Surprise	0.0049	0.0068	0.0047	0.0051	0.0039	0.0045
Trust	0.0393	0.0380	0.0386	0.0356	0.0371	0.0414

To summarise, the emotion trends were consistent across both datasets, with the {Trust, Anticipation, Joy} set accounting for the dominant emotions and {Disgust, Fear} representing the least dominant emotions. Trust emerged as the highest emotion, followed by anticipation and joy, while disgust was the least prevalent emotion, followed by anger.

From Table 6, the year 2020 had the highest emotion scores for disgust, fear, anger, and negative valence. It topped all negative emotions, including negative valence, from 2012 onwards. As mentioned above, the outbreak of COVID-19 might be the reason

for the highest negative emotion scores in 2020. On the other hand, 2020 had the lowest trust and positive valence scores, where trust is the dominant emotion and positive is the dominant valence across all years from 2012. In contrast, 2019 had the highest trust, the dominant emotion across all years, and the highest joy scores. Additionally, 2013 had the highest anticipation, surprise, and positive valence scores, as well as the lowest fear scores across all years from 2012 to 2022.

Table 6 Highest and lowest emotions from 2012-2022

	Trust	**Anticipation**	**Joy**	**Disgust**	**Fear**	**Sadness**	**Surprise**	**Anger**	**Positive**	**Negative**
Highest	2019	2013	2019	**2020**	**2020**, 2022	2016	2013	2016 2020 2022	2013	**2020**
Lowest	**2020** 2015	2021	2014	2012	2013	2017	**2020**	2012	**2020**	2012

The improved writing provides a clear overview of the dominant emotions across years, highlights the consistent trends observed in both datasets, and emphasises the impact of COVID-19 on emotion scores in 2020. It also includes specific details about the highest and lowest values for various emotions and valences, allowing for a comprehensive understanding of the emotion distribution across the years.

5.2 Hybrid machine learning approach

Three different experiments were conducted using a hybrid machine learning approach, i.e., one experiment using the 2018-2022 dataset, another experiment using the 2012-2022 dataset, and the last experiment with balanced class weights using the 2012-2022 dataset. The two datasets annotated with dominant emotions from the NRC Emotion Lexicon were given as input to Word2Vec, which generated 218-word embeddings for the 218 annotated records in the 2018-2022 dataset and 718 word embeddings in the 2012-2022 dataset, respectively.

Word2Vec generates vectors out of words or word embeddings, treating similar meaning words as the same and assigning them the same vector, while assigning different vectors to words with different meanings. It calculates the cosine (cos) similarity for two words; if the cos value is "1," both words are assigned the same vector. The word embeddings generated for the dataset with 218 records had a dimension of 300 x 218, meaning Word2Vec used a vector space of 300 coordinates for a single abstract, and there were 218 abstracts in the 2018-2022 datasets. A pickle

file containing the word embeddings was generated and later used to train the Random Forest classifier. The abstracts were split into an 80:20 ratio, with 80% of the records used to train the model and 20% for testing. The prediction model built using the 2018 to 2022 annotated records yielded an accuracy of 0.68 in experiment 3, and other performance metrics, such as precision, recall, and F1-score, were calculated.

The same model was trained and tested with the 2012-2022 dataset too, where abstracts from 2012 to 2022, total 718 abstracts, were merged into a single training dataset. In this scenario, Word2Vec generated 718-word embeddings, which were used to train the Random Forest classifier. The prediction model trained with the larger dataset achieved an accuracy of 0.60, a decrease of 8% compared to the previous experiment.

The emotion labels in the dataset were trust, anticipation, positive, negative, and fear. The majority of records were annotated with positive, trust, and anticipation, while very few were labelled as negative and fear, as scientific papers typically contain formal and positive words unlike text on other online platforms or social media. This created a significant imbalance in the dataset, as the distribution of records with respect to emotions was not equal.

To attain balance in the emotion classes, the "class_weight" function of the Random Forest classifier was assigned "balance" to balance the dataset, and the model was trained in the last experiment. The model achieved an accuracy of 0.59 in the last experiment, which was less than the accuracy of the model trained with the imbalanced dataset in the previous experiment. The difference in accuracies was just one percent, where the performance metrics of precision, recall, and F1-scores were almost the same in both experiments.

The accuracy of the model trained with the balanced dataset was 9% less than the model trained with the 2018-2022 dataset, which was not balanced in experiment 3. Even though the size of the dataset was increased to 718 records by including all abstracts from 2012 to 2022, it is still relatively small to train a machine learning model effectively. Usually, the size of a training dataset required to properly train a model is much larger, almost double the size of the increased dataset from 2012 to 2022, and the labels should be evenly distributed to train the model in all possible ways.

In general, the dataset of abstracts from PRO-VE is not evenly distributed because scientific paper abstracts are simple and formal text, which do not contain emotion-rich words to extract all eight emotions from the abstracts. Additionally, the dataset size is small, making it challenging to train a robust machine learning model effectively.

6 Conclusion

This research commenced with a comprehensive background study on sentiment analysis, emotion detection applications, and existing related work and theories on emotions. Two emotion detect techniques, a lexicon-based approach and a hybrid machine learning-based approach, were designed and employed to detect emotions in scientific texts.

Through a series of experiments, dominant emotions were identified in the datasets by calculating average emotion scores across all years. Trends and patterns were

analysed based on the experimental findings, revealing insights into the emotional landscape of the PRO-VE conference abstracts.

Additionally, a hybrid machine learning model was developed using Word2Vec and the Random Forest classifier, trained on a labelled dataset to predict emotions. The performance of this hybrid approach was thoroughly analysed, contributing to the understanding of emotion detection in scientific texts.

While detecting emotions from text is a complex process due to the multiple meanings of words and the context-dependent nature of language, this research tackled the unique challenges posed by scientific papers. Unlike datasets from human-computer interaction, social media analysis, services industry, and customer management, which often contain emotion-rich words, the PRO-VE conference abstracts exhibited a predominance of positive emotions and a higher positive valence score. This can be attributed to the nature of scientific writing, where positive emotions like trust, anticipation, and joy are more commonly conveyed.

It is important to acknowledge the limitations of this research, which focused solely on the PRO-VE conference, and the dataset created was annotated primarily with positive emotions conveyed by the authors through their abstracts. Furthermore, the dataset size of 718 records, while substantial, may still be insufficient to train a highly accurate machine learning model, as the creation of larger datasets requires significant time and effort.

Looking ahead, future research directions could involve enhancing the dataset quality by incorporating additional fields related to authors and their backgrounds, enabling investigations into the relationship between the intensity of emotions and the number or background of authors. Including indexed keywords and author-provided keywords could also facilitate the analysis of how these keywords are associated with the conveyed emotions.

Increasing the dataset size by incorporating historical conference abstracts from the PRO-VE series dating back to 1999 and extending the analysis to other conferences, such as "CAiSE," which has 35 conference series entirely focused on computer science, could further enrich the dataset and provide more comprehensive insights.

Moreover, exploring alternative emotion classification schemes beyond Plutchik's eight-wheel of emotions could lead to the identification of emotion classes better suited for capturing the nuances of emotions expressed in scientific and computer science fields.

By addressing these future research directions, including expanding the dataset with additional author-related information, increasing the dataset size through historical conference abstracts (or adding other conferences), and exploring alternative emotion classification frameworks, future studies can potentially provide more comprehensive insights into the emotional aspects of scientific writing and enhance the understanding of emotional disclosure in academic research, particularly in the fields of computer science, information systems and related disciplines.

References

1. Camarinha-Matos LM, Afsarmanesh H (2005) Collaborative networks: A new scientific discipline. J Intell Manuf 16:439–452.
2. Ferrada F, Camarinha-Matos LM (2017) A system dynamics and agent-based approach to model emotions in collaborative networks. IFIP Adv Inf Commun Technol 499:29–43.
3. Lerner JS, Li Y, Valdesolo P, Kassam KS (2015) Emotion and decision making. Annu Rev Psychol 66:799–823.
4. Chatterjee A, Gupta U, Chinnakotla MK, et al (2019) Understanding Emotions in Text Using Deep Learning and Big Data. Comput Human Behav 93:309–317.
5. Marín-Morales J, Llinares C, Guixeres J, Alcañiz M (2020) Emotion Recognition in Immersive Virtual Reality: From Statistics to Affective Computing.
6. Ekman P, Levenson RW, Friesen W V. (1983) Autonomic Nervous System Activity Distinguishes Among Emotions. Science (1979) 221:1208–1210.
7. PLUTCHIK R (1980) A GENERAL PSYCHOEVOLUTIONARY THEORY OF EMOTION. Theories of Emotion 3–33.
8. Sailunaz K, Alhajj R (2019) Emotion and sentiment analysis from Twitter text. J Comput Sci 36:101003.
9. Mohammad SM, Turney PD (2013) CROWDSOURCING A WORD-EMOTION ASSOCIATION LEXICON. Comput Intell 29:.
10. Wang W, Chen L, Thirunarayan K, Sheth AP (2012) Harnessing twitter "big data" for automatic emotion identification. Proceedings - 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust and 2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012 587–592.
11. Gievska S, Koroveshovski K, Chavdarova T (2015) A hybrid approach for emotion detection in support of affective interaction. IEEE International Conference on Data Mining Workshops, ICDMW 2015-January:352–359.
12. Sarsam SM, Al-Samarraie H, Alzahrani AI, et al (2021) A lexicon-based approach to detecting suicide-related messages on Twitter. Biomed Signal Process Control 65:102355.
13. Park SH, Bae BC, Cheong YG (2020) Emotion recognition from text stories using an emotion embedding model. Proceedings - 2020 IEEE International Conference on Big Data and Smart Computing, BigComp 2020 579–583.
14. Pouransari H, Ghili S (2014) Deep learning for sentiment analysis of movie reviews. CS224N Proj 1–8
15. Deho OB, Agangiba WA, Aryeh FL, Ansah JA (2018) Sentiment analysis with word embedding. IEEE International Conference on Adaptive Science and Technology, ICAST 2018-August: