(Online) ISSN 2303-4521
Periodicals of Engineering and Natural Sciences *Original Research*
Vol. 13, No. 2, June 2025, pp.435-444

Sentiment analysis in Arabic panic detection systems in Iraq

Sameerah Faris Khlebus[1], Mohammed Salih Mahdi[2], Monji Kherallah[3], Ali Douik[4]
[1] National School of Electronics and Communications in Sfax, University of Sfax, Tunisia
[2] Business Information College, University of Information Technology and Communication, Baghdad, Iraq
[3] Faculty of Sciences of Sfax, University of Sfax, Tunisia
[4]University of Sousse, Sousse, Tunisia

*Corresponding author E-mail: sameerah.alradhi@uoitc.edu.iq

ABSTRACT

The internet's prevalence has increased the use of social media platforms for sharing peoples' views, reviews, opinions, and sentiments. Consequently, sentiment analysis has become valuable to marketing, brand management, public policy, healthcare, and e-commerce. Panic detection systems refer to a system that utilizes sentiment analysis techniques to identify, quantify, and potentially track expressions of panic, fear, or related strong negative emotions within digital content. Sentiment analysis application can indicate a person's attitude or cultural influences. The rise of Arabic digital content on social platforms presents unique challenges and opportunities for sentiment analysis. There are fewer Arabic sentiment lexicons, emotional connotations, and annotated corpora than in English. Therefore, this study applies machine learning and deep learning to identify sentiments within the IRAQI-Arabic panic detection systems. The proposed model is based on Residual Network (ResNet), MobileNet, and Convolutional Neural Network (CNN). A web scraping API collected COVID-19-related news from Google, Facebook, Twitter, and BBC News. The model used was retrained on the images and comments based on training, public, and private sub-sets. Bidirectional Encoder Representations from Transformers (BERT) were applied across SMOTE, Random oversampling, five-folds, and ten folds. The results show that the accuracy of BERT is 91%. The proposed MobileNet model had an accuracy of 0.98, recall of 0.98, and MAE of 0.032. The model's performance confirms that it is suitable for critical applications that require high-level precision. The study offers a novel model for Arabic panic systems. Also, it presents an Iraqi-Arabic dataset tailored for panic detection.

Keywords: Sentimental analysis, Panic detection system, Artificial intelligence, Deep learning, Arab tweets

1. Introduction

The internet's pervasiveness has led to the rise of social media and blog platforms, which have become powerful sources of online news and opinions. This has significantly increased the number of people sharing views, reviews, comments, and ratings on various products and services [1]. Digital communication's ubiquity has identified sentiment analysis as an integral component of natural language processing (NLP). Sentiment analysis, or opinion mining, discerns and categorizes the emotional tone within textual data [2]. It aims to provide insights into public opinions, products, and societal attitudes [3, 4]. Furthermore, it supports decision-making in marketing, government, product development, healthcare, media, education, and e-commerce [5, 6]. Individuals' views, emotions, feelings, and opinions were important information for decision-making before the widespread of the World Wide Web [7]. Sentiment analysis application can indicate a person's attitude or cultural influences.

Recently, social platforms experienced an increase in the use of sentiment, which drew the attention of many researchers due to its importance in vast areas of our daily activities. Similarly, there has been significant growth in online Arabic content [8, 9]. The rise of Arabic digital content on social platforms presents unique challenges and opportunities for sentiment analysis. Dissimilar to English, Arabic has fewer sentiment lexicons, emotional connotations, and annotated corpora [10]. Researchers have applied sentiment analysis to Arabic [11, 12]. They are constantly developing resources to improve the accuracy and applicability of Arabic sentiment analysis to understand public opinion, consumer attitudes, and trends within Arabic-speaking communities [8]. Nevertheless, there is little available literature on sentimental analysis for Arabic text [13, 14]. This limitation could be due to the complexity of Arabic's complex characteristics, such as its semantics, grammar, dialects, and spellings. Furthermore, Arabic can be classed into Modern Standard Arabic (MSA), dialect Arabic (DA), and classical Arabic [10]. Therefore, this study applies machine learning and deep learning to identify sentiments within the IRAQI-Arabic panic detection systems. In this context, the system is designed to discern expressions of panic that are likely triggered by news, social media discussions, and public discourse surrounding a crisis.

Recently, machine-learning approaches have been extensively used for sentiment analysis. [15] observed that supervised algorithms perform better on training data and with low-performance accuracy on the test data. [14] found that 18 Arabic COVID-19 Apps received positive reviews from 114,499 users. [13] performed an Arabic sentiment analysis of 2,247 tweets using machine learning classification algorithms. Their findings showed that SVM exhibited a 90% accuracy. Similarly, [16] applied Naïve Bayes (NB) and SVM to assess customer retention on Instagram. The analysis revealed that SVM had a higher accuracy of 76.59%. A study on sentiment analysis for customers of a water company in Indonesia showed that SVM had the highest accuracy of 95% [17].

Some studies have focused on sentiment analysis of Arabic text. [18] Arabic sentiment analysis involved 1000 datasets from Facebook and Twitter using DT and SVM. Their findings showed that SentiStrength had a 95.59% accuracy. Nevertheless, the authors used limited datasets, which affected the generalizability of their findings. [19] applied deep learning techniques to sentiment analysis for 1098 Arabic tweets. They aim to determine the sentiments associated with specific topics from Arabic social media content. However, the classification algorithms were limited. Results suggest that the Neural Networks classifier with Word Embedding achieved 81% accuracy. [20] explored supervised machine learning in Arabic sentiment classification on the document level. The study findings indicated that SVM and NB were effective. However, a limited number of algorithms were used in their study. [9] applied sentiment analysis to examine 10,646 data using machine learning approaches. The results showed that support vector machines and decision trees had a higher accuracy of 95%.

Many papers employing deep learning (DL) algorithms were evaluated for Arabic text. For example, the study of [21] investigated these models for opinion classification on the sentence level. The findings confirm Recursive Auto Encoders as an effective model for classifying sentence polarity. According to [22], SVM is the most suitable for sentiment analysis of Arabic text based on an accuracy of 83%. The authors in [23] applied sentiment analysis to 815 comments from an e-newspaper in Saudi Arabia using a hybrid NB classifier. The system achieved an accuracy of 82% across four categories. The authors in [24] developed an Arabic Twitter Sentiment Analysis (ATSA) model based on supervised ML. They evaluated the model using NB and SVM in four representation approaches. SVM showed the highest accuracy of 91.15%. Regardless, the representations discarded some distinctive word features identifying the sentiment class.

The findings of this study aim to uncover the nuances of Arabic sentiment and enhance its applicability, particularly in the context of panic detection systems. By evaluating the comparative classification performance of various sentiment analysis techniques on Arabic panic systems, the research sheds new light on this area and introduces a novel model specifically designed for such applications. Additionally, the study presents an Iraqi-Arabic dataset tailored for panic detection and highlights the potential of Random-BERT as a reliable approach for sentiment analysis across diverse applications. The paper is structured into several sections: Section 2 outlines the methodology, Section 3 summarizes the findings, Section 4 discusses these findings, and Section 5 concludes with implications for both practice and research.

2. Methodology

The proposed model is based on Residual Network (ResNet), MobileNet, and convolutional neural network (CNN). Deep models like CNNs (MobileNet, ResNet) and BERT have gained preference over traditional machine learning due to its ability to automatically extract features, handle complex and non-linear relationships in large and diverse datasets [19]. Additionally, they usually achieve higher accuracy in image recognition and natural language processing. This approach ensures gathering real-time and raw data to capture prevalent online sentiments effectively. The model integrates algorithmic optimization, parallel processing methods, and cutting-edge deep learning technologies to achieve low latency and high reliability. The performance metrics used are accuracy, precision, and recall [25, 26]. The SMOTE-BERT experiment uses five-fold and ten-fold cross-validation. The increase in folds enables a more robust evaluation of the model's performance, leading to higher average metrics.

2.1. Data scraping

A web scraping API collected COVID-19-related news from Google, Facebook, Twitter, and BBC News [27]. Web crawling is used to identify relevant web pages, while scraping gathers relevant data from those pages [27]. The parameters used in the API included the title (“COVID,” “panic,” “diseases,” “crises,” “attack”) and locality

("Iraq"). This enabled the collection of news articles, posts, and tweets related to panic developments in Iraq. Headlines, summaries, and publication dates of relevant news articles were collected from Google. On Facebook, Iraqi public posts, articles, and shared content on the platform were accessed. Concerning Twitter, tweets, timestamps, user information, and engagement metrics were retrieved. Meanwhile, Iraq-related article headlines, summaries, publication dates, and source information were extracted from BBC News. Finally, duplicates were removed, and the remaining data was standardized.

Iraqi-based sentimental-based contexts related to panic were distributed in 2524 rows and classified into panic, non-panic, and neutral for accuracy and reliability. The dataset is annotated based on the sentiment conveyed in each new instance. The panic-related (and non-panic and neutral) posts were labeled through manual annotation The text is preprocessed to remove noise and irrelevant information. This process ensures that the analysis format is consistent for the sentiment analysis.

Data augmentation via an oversampling strategy is employed to address potential imbalances in the datasets. Imbalanced datasets can lead to biased model training. In this study, the Synthetic Minority Over-Sampling Technique (SMOTE) and Random Oversampling are used to prevent the model from being biased toward the majority class. These approaches were applied to increase the accuracy of the proposed model and enhance its prediction capability.

The model's performance is refined to achieve suitable metrics in the training optimization process for sentiment detection from textual data. This iterative process involves several key steps to enhance the models' effectiveness. The models are exposed to the training data over multiple iterations or epochs during training. Accuracy, precision, recall, and F1-scores are computed to assess the models' effectiveness in sentiment detection. Hyperparameter tuning is employed to adjust the learning rate, batch size, and regularization strength. Dropout and L2 regularization are also applied to prevent overfitting and improve the model's generalization ability. This approach mitigates the risk of the models memorizing the training data and performing poorly on unseen data. For the textual data, Bidirectional Encoder Representations from Transformers (BERT), Multilingual D-BERT, Cross-lingual Language Model BERT, and Generative Pre-trained Transformer (GPT) to achieve the best performance for natural language processing techniques [28, 29].

The dataset images are labeled iteratively for accurate and consistent annotation. Hence, specific attributes, categories, and characteristics are labeled within the images. The annotators manually review the images and apply the appropriate labels based on the established guidelines. Quality control involves reviewing the subset of annotated images for errors or discrepancies. This study employs the Facial Expression Recognition 2013 (FER2013) dataset comprising 35,887 grayscales. The images are categorized into anger, disgust, fear, happiness, sadness, surprise, and neutral. The seven categories are depicted in Figure 1. Furthermore, the images are divided into training, pubic, and private sub-sets.

Figure 1. Example of illustrating six emotional states

3. Results

Table 1 shows the SMOTE-BERT and Random-BERT experiments across five-fold cross-validation. SMOTE-BERT's accuracy range from 0.75 to 0.82, precision from 0.78 to 0.85, and recall between 0.72 to 0.78. Experiment 4 had the highest accuracy (0.82), precision (0.85), and recall (0.78). Meanwhile, experiment 2 exhibits higher precision and recall than others. Concerning Random-BERT, the accuracy ranges from 0.76 to 0.80, precision from 0.79 to 0.82, and recall from 0.75 to 0.79. Experiment 4 also showed correspondingly the highest accuracy, precision, and recall of 0.80, 0.82, and 0.79. This finding indicates that the configuration used in Experiment 4 may be more effective for sentiment analysis.

Table 1. Five folds results of SMOTE-BERT and Random-BERT methods

Experiment	SMOTE-BERT			Random-BERT		
	Accura cy	Precisio n	Recall	Accura cy	Precisio n	Recall
Experiment 1	0.78	0.80	0.75	0.79	0.80	0.77
Experiment 2	0.80	0.82	0.76	0.78	0.82	0.76
Experiment 3	0.75	0.78	0.72	0.77	0.81	0.78
Experiment 4	0.82	0.85	0.78	0.80	0.82	0.79
Experiment 5	0.76	0.79	0.74	0.76	0.79	0.75

The ten-fold cross-validation for SMOTE-BERT and Random-BERT is illustrated in Table 2 In SMOTE-BERT, experiments 4 and 10 had the highest accuracy, precision, and recall. For Random-BERT, it ranges from 0.87 to 0.91, precision from 0.90 to 0.93, and recalls from 0.86 to 0.90. The accuracy was between 0.75 to 0.81, precision from 0.79 to 0.85, and recall from 0.72 to 0.77. This finding suggests a more significant variability in performance due to the larger number of folds used for cross-validation.

Table 2. Ten folds results of SMOTE-BERT and Random-BERT methods

Experiment	SMOTE-BERT			Random-BERT		
	Accur acy	Precis ion	Recal l	Accur acy	Precis ion	Recal l
Experiment 1	0.82	0.85	0.80	0.80	0.82	0.76
Experiment 2	0.81	0.84	0.79	0.78	0.80	0.75
Experiment 3	0.80	0.82	0.78	0.81	0.84	0.77
Experiment 4	0.83	0.87	0.81	0.77	0.81	0.74
Experiment 5	0.84	0.88	0.82	0.79	0.83	0.76
Experiment 6	0.79	0.83	0.77	0.75	0.79	0.72
Experiment 7	0.82	0.86	0.8	0.81	0.86	0.77
Experiment 8	0.81	0.84	0.79	0.76	0.80	0.73
Experiment 9	0.80	0.85	0.78	0.80	0.85	0.76
Experiment 10	0.83	0.88	0.81	0.78	0.82	0.75

The five-fold results of SMOTE D-BERT and Random D-BERT are presented in Table 3. SMOTE D-BERT's accuracy ranges from 0.85 to 0.90, precision 0.88 to 0.92, and recall from 0.83 to 0.89. These results indicate that the SMOTE D-BERT model achieves high metrics across different configurations. The experiment demonstrated the highest accuracy, precision, and recall of 0.90, 0.92, and 0.89, respectively. Random-DBERT accuracy was from 0.87 to 0.91, precision from 0.90 to 0.93, and recall from 0.86 to 0.90. Hence, Random-D-BERT has high accuracy, precision, and recall across different configurations. Experiment 3 had the highest accuracy, precision, and recall of 0.91, 0.93, and 0.90, respectively.

Table 3. Five folds results of SMOTE D-BERT and Random D-BERT methods

Experiment	SMOTE D-BERT			Random D-BERT		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Experiment 1	0.88	0.90	0.86	0.79	0.8	0.77
Experiment 2	0.85	0.88	0.83	0.78	0.82	0.76
Experiment 3	0.90	0.92	0.89	0.77	0.81	0.78
Experiment 4	0.86	0.89	0.85	0.80	0.82	0.79
Experiment 5	0.89	0.91	0.87	0.76	0.79	0.75

Table 4 highlights the ten-fold results of SMOTE D-BERT and Random D-BERT. SMOTE D-BRET accuracy ranges from 0.86 to 0.93, precision from 0.88 to 0.95, and recalls from 0.85 to 0.92. Experiment 8 showed the highest accuracy, precision, and recall of 0.93, 0.95, and 0.92, respectively. For Random D-BERT, accuracy ranges from 0.90 to 0.95, precision from 0.92 to 0.96, and recall from 0.89 to 0.92. Similarly, experiment 8 displayed higher accuracy, precision, and recall, achieving values of 0.95, 0.96, and 0.92, respectively.

Table 4. Ten- folds results of SMOTE-DBERT and Random D-BERT methods

Experiment	SMOTE D-BERT			Random D-BERT		
	Accu racy	Preci sion	Recal l	Accu racy	Preci sion	Recal l
Experiment 1	0.91	0.93	0.90	0.80	0.82	0.76
Experiment 2	0.92	0.94	0.91	0.78	0.80	0.75
Experiment 3	0.93	0.95	0.92	0.81	0.84	0.77
Experiment 4	0.90	0.92	0.89	0.77	0.81	0.74
Experiment 5	0.94	0.95	0.91	0.79	0.83	0.76
Experiment 6	0.92	0.94	0.90	0.75	0.79	0.72
Experiment 7	0.91	0.93	0.91	0.81	0.86	0.77
Experiment 8	0.95	0.96	0.92	0.76	0.8	0.73
Experiment 9	0.93	0.95	0.91	0.80	0.85	0.76
Experiment 10	0.92	0.94	0.90	0.78	0.82	0.75

SMOTE-GPT and Random-GPT results across five folds are presented in Table V. SMOTE-GPT accuracy ranges from 0.72 to 0.78, precision from 0.76 to 0.80, and recall from 0.71 to 0.76. Meanwhile, experiments 2 and 5 performed better, with an accuracy of 0.78. Random-GPT demonstrated accuracy ranging from 0.73 to 0.77, precision from 0.76 to 0.79, and recall from 0.70 to 0.75. Experiment 3 had a higher precision of 0.75 and recall of 0.77.

Table 5. Five folds results of SMOTE-GPT and Random-GPT methods

Experiment	SMOTE-GPT			Random-GPT		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Experiment 1	0.75	0.77	0.74	0.77	0.71	0.74
Experiment 2	0.78	0.79	0.76	0.78	0.72	0.76
Experiment 3	0.72	0.76	0.77	0.79	0.75	0.77
Experiment 4	0.76	0.78	0.73	0.76	0.70	0.73
Experiment 5	0.78	0.80	0.75	0.79	0.74	0.75

Table 6 highlights the SMOTE-GPT and Random-GPT results based on ten-fold. The accuracy of SMOTE-GPT ranges from 0.72 to 0.79, precision is from 0.78 to 0.82, and recall is from 0.71 to 0.77. Experiment 5 had higher accuracy, precision, and recall of 0.79, 0.82, and 0.77. Concerning Random-GPT, the accuracy ranges from 0.70 to 0.77, precision from 0.76 to 0.81, and recall from 0.69 to 0.74. Experiment 4 had a higher accuracy of 0.81, a precision of 0.74, and recall of 0.76.

Table 6. Ten folds results of SMOTE-GPT and Random-GPT methods

Experiment	SMOTE-GPT			Random-GPT		
	Acc urac y	Prec isio n	Rec all	Acc urac y	Prec isio n	Rec all
Experiment 1	0.76	0.79	0.74	0.77	0.72	0.74
Experiment 2	0.78	0.81	0.70	0.78	0.73	0.70
Experiment 3	0.77	0.80	0.72	0.79	0.71	0.72
Experiment 4	0.75	0.79	0.76	0.81	0.74	0.76
Experiment 5	0.79	0.82	0.71	0.80	0.72	0.71
Experiment 6	0.72	0.78	0.73	0.80	0.70	0.73
Experiment 7	0.74	0.8	0.71	0.76	0.69	0.71
Experiment 8	0.77	0.81	0.77	0.79	0.74	0.77
Experiment 9	0.76	0.8	0.74	0.81	0.73	0.74
Experiment 10	0.75	0.78	0.76	0.80	0.74	0.76

Results of SMOTE-XLM-BERT and Random- XLM-BERT are depicted in Table 7. SMOTE-XML-BRET model demonstrates an accuracy from 0.84 to 0.88, precision from 0.87 to 0.91, and recall from 0.83 to 0.87. Experiment 2 has a higher performance with an accuracy of 0.88. For Random- XLM-BERT, the accuracy ranges from 0.86 to 0.92, precision from 0.89 to 0.94, and recall from 0.85 to 0.90. Experiment 2 had a higher accuracy of 0.88, a precision of 0.91, and recall of 0.87.

Table 7. Five folds results of SMOTE-XLM-BERT and Random-XLM-BERT methods

Experiment	SMOTE-XLM-BERT			Random-XML-BERT		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Experiment 1	0.85	0.87	0.83	0.85	0.88	0.84
Experiment 2	0.88	0.91	0.87	0.88	0.90	0.86
Experiment 3	0.86	0.89	0.85	0.89	0.92	0.88
Experiment 4	0.87	0.91	0.87	0.86	0.89	0.85
Experiment 5	0.84	0.88	0.84	0.87	0.91	0.87

As in Table 8, SMOTE-XLM-BERT demonstrated an accuracy from 0.86 to 0.92, precision from 0.89 to 0.94, and recall from 0.85 to 0.90. Experiment 7 showed a 0.82 accuracy, 0.86 precision, and 0.80 recall. Moreover, Random-XML-BERT accuracy ranges from 0.89 to 0.93, precision from 0.92 to 0.95, and recalls from 0.87 to 0.91. Experiment 3 exhibited an accuracy of 0.80, precision of 0.82, and recall of 0.78.

Table 8. Ten folds results of SMOTE-XLM-BERT and Random XLM-BERT methods

Experiment	SMOTE-XLM-BERT			Random-XML-BERT		
	Accu racy	Preci sion	Reca ll	Accu racy	Preci sion	Reca ll
Experiment 1	0.88	0.90	0.86	0.90	0.92	0.89
Experiment 2	0.90	0.92	0.89	0.91	0.94	0.90
Experiment 3	0.89	0.91	0.87	0.92	0.95	0.91
Experiment 4	0.91	0.94	0.90	0.89	0.93	0.88
Experiment 5	0.87	0.92	0.88	0.93	0.95	0.90
Experiment 6	0.86	0.89	0.85	0.91	0.92	0.89
Experiment 7	0.92	0.93	0.88	0.90	0.95	0.88
Experiment 8	0.89	0.94	0.89	0.90	0.93	0.89
Experiment 9	0.91	0.93	0.90	0.89	0.94	0.90
Experiment 10	0.88	0.91	0.86	0.89	0.92	0.87

The accuracies of MobileNet, ResNet, and CNN are determined using 10-fold cross-validation in Table IX. The results are displayed in Table 9. According to the Table, MobileNet had the highest accuracy of 0.98 and recall of 0.98. Similarly, MobileNet had the lowest MAE of 0.032.

Table 9. Metrics for the 3 models

	Accuracy	Recall	MAE
MobileNet	0.980	0.969	0.032
ResNet	0.930	0.919	0.071
CNN	0.900	0.880	0.100

This study explored sentiment analysis for Arabic panic detection systems. A comparative analysis of the four approaches reveals BERT's superiority in sentiment classification. Across SMOTE, Random oversampling, five-folds, and ten-folds, BERT showed a 91% accuracy. The result aligns with previous studies [30, 31, 32]. Wang et al. [33] also demonstrated that BERT has a high accuracy. This outcome is likely due to BERT's ability to learn context-sensitive word representations. Additionally, it can effectively handle complex syntactic and semantic relationships. Furthermore, BERT is considered a better training methodology with augmentation techniques and a larger dataset [34].

Furthermore, the five- and ten-fold cross-validation results highlight SMOTE-D-BERT’s effectiveness in sentiment analysis tasks. These findings highlight the model’s consistently high performance and robustness across different configurations and datasets. Hence, it highlights its potential as a reliable approach for sentiment analysis in diverse applications. Although the SMOTE-GPT model demonstrates moderate performance, there is variability in performance across different configurations. Therefore, its performance could be based on s different configurations or datasets.

The data showed that MobileNet performed better than ResNet and CNN. MobileNet had an accuracy of 0.98. Previous studies also demonstrated MobileNet’s higher accuracy ability [35, 36]. The findings suggest that

MobileNet can detect panic effectively through image-based inputs with fewer errors. Therefore, the model's recall was 0.98 and an MAE of 0.032.

Similarly, studies showed that MobileNet had a high recall rate in detecting pneumonia from chest X-ray images [37]. The lower MAE shows a very low average deviation between the predicted and actual sentiment values. MobilNet's low MAE and high accuracy indicate that the model is consistent and reliable across different inputs. This means that the model can extract fine-grained features in Arabic text effectively. Thus, the complexity of the Arabic language and different dialects will not affect MobileNet's performance. Finally, MobileNet model outperforms all the tested BERT-based models based on accuracy and recall, achieving significantly higher scores (0.98 vs. a maximum of around 0.93-0.95) across different test conditions.

4. Conclusion

This study applies machine learning and deep learning to identify sentiments within the IRAQI-Arabic panic detection systems. The proposed model is based on Residual Network (ResNet), MobileNet, and Convolutional Neural Network (CNN). The four models including MobileNet, ResNet, and CNN, were assessed to identify panic detection through deep learning. MobileNet emerged as the most effective model for image-based panic detection across five and ten folds. The model's performance confirms its suitability for critical applications requiring high-level precision. Hence, it can effectively capture the nuances of the Arabic language's complexity, dialects, and varied expressions. Overall, the results highlight Random-BERT's effectiveness for sentiment analysis tasks. This technique consistently demonstrated high accuracy, precision, and recall. These findings emphasize the potential of Random-BERT as a reliable approach for sentiment analysis in diverse applications.

The study sheds new light on the comparative classification performance evaluation of various sentiment analysis techniques on Arabic panic systems. One BERT limit is that it requires computational solid capacity. Also, the limited dataset in this study could cause overfitting, reducing its ability to generalize to new data. Despite the limitations, insights from this study can improve Arab sentiment classification based on this technique. Previous studies have emphasized the use of large datasets, which require more computational resources. Contrastingly, this study focused on effectively identifying sentiments in real-world Arab panic detection systems.

Declaration of competing interest

The authors declare that they have no known financial or non-financial competing interests in any material discussed in this paper.

Funding information

The authors declare that they have received no funding from any financial organization to conduct this research.

Author contribution

Sameerah Faris Khlebus: Conceptualized the research idea, conducted the literature review, and contributed to the methodology design. He was also responsible for data collection and analysis, and drafted the initial version of the manuscript.

Mohammed Salih Mahdi: Provided guidance on the research framework, contributed to the interpretation of results, and assisted in refining the manuscript's structure and content. He also played a key role in reviewing and editing the final manuscript.

Monji Kherallah: Contributed to the development of the theoretical framework and provided insights into the technological aspects of the study. He assisted with data analysis and contributed to writing specific sections of the manuscript.

Ali Douik: Provided critical feedback on the research design and methodology. He contributed to the discussion of findings and helped in revising the manuscript for clarity and coherence.

All authors have read and approved the final manuscript.

References

[1] P. Sahu, S. K. Mohapatra, P. K. Sarangi, J. Mohanty, and P. K. Sarangi, “Detection of non-melanoma skin cancer by deep convolutional neural network and stochastic gradient descent optimization algorithm,” J. Mech. Cont. Math. Sci., vol. 20, no. 1, pp. 1–15, Jan. 2025. https://doi.org/10.26782/jmcms.2025.01.00005.

[2] H. Singh, R. Tripathy, P. K. Sarangi, U. Giri, S. K. Mohapatra, and N. R. Amin, “Feature-based implementation of machine learning algorithms for cardiovascular disease prediction,” J. Mech. Cont. Math. Sci., vol. 19, no. 11, pp. 1–15, Nov. 2024. https://doi.org/10.26782/jmcms.2024.11.00008.

[3] C. Puschmann and A. Powell, “Turning Words Into Consumer Preferences: How Sentiment Analysis Is Framed in Research and the News Media,” *Social Media and Society*, vol. 4, 2018. https://doi.org/10.1177/2056305118797724

[4] N. A. Sharma, A. B. M. S. Ali, and M. A. Kabir, “A review of sentiment analysis: tasks, applications, and deep learning techniques,” 2024.

[5] G. E. S. Contreras and I. Aguilar-Alonso, “Influence of the Application of Sentiment Analysis as a Process Improvement in Different Sectors,” in *ITMS 2021 - 62nd Int. Sci. Conf. on Information Technology and Management Science*, 2021.

[6] N. B. V. Le and J. H. Huh, “Applying sentiment product reviews and visualization for BI systems in Vietnamese e-commerce website: Focusing on Vietnamese context,” *Electronics (Switzerland)*, vol. 10, 2021. https://doi.org/10.3390/electronics10202481

[7] M. Tareq *et al.*, “Data-Augmentation for Bangla-English Code-Mixed Sentiment Analysis: Enhancing Cross Linguistic Contextual Understanding,” *IEEE Access*, vol. 11, pp. 51657–51671, 2023. https://doi.org/10.1109/ACCESS.2023.3277787

[8] S. O. Alhumoud and A. A. Al Wazrah, “Arabic sentiment analysis using recurrent neural networks: a review,” *Artif. Intell. Rev.*, vol. 55, pp. 707–748, 2022. https://doi.org/10.1007/s10462-021-09989-9

[9] O. Alsemaree, A. S. Alam, S. S. Gill, and S. Uhlig, “Sentiment analysis of Arabic social media texts: A machine learning approach to deciphering customer perceptions,” *Heliyon*, vol. 10, 2024. https://doi.org/10.1016/j.heliyon.2024.e27863

[10] H. Elfaik and E. H. Nfaoui, “Deep Bidirectional LSTM Network Learning-Based Sentiment Analysis for Arabic Text,” *J. Intell. Syst.*, vol. 30, pp. 395–412, 2021. https://doi.org/10.1515/jisys-2020-0021

[11] A. Al-Hashedi *et al.*, “Ensemble Classifiers for Arabic Sentiment Analysis of Social Network (Twitter Data) towards COVID-19-Related Conspiracy Theories,” *Appl. Comput. Intell. Soft Comput.*, vol. 2022, 2022. https://doi.org/10.1155/2022/6614730

[12] S. R. El-Beltagy and A. Ali, “Open issues in the sentiment analysis of Arabic social media: A case study,” in *9th Int. Conf. on Innovations in Information Technology*, pp. 215–220, 2013.

[13] A. Nasayreh *et al.*, “Arabic Sentiment Analysis for ChatGPT Using Machine Learning Classification Algorithms,” *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 23, 2024. https://doi.org/10.1145/3638285

[14] M. Ramzy and B. Ibrahim, “User satisfaction with Arabic COVID-19 apps: Sentiment analysis of users’ reviews using machine learning techniques,” *Inf. Process. Manag.*, vol. 61, 2024. https://doi.org/10.1016/j.ipm.2024.103644

[15] D. M. Witten, “Classification and clustering of sequencing data using a poisson model,” *Ann. Appl. Stat.*, vol. 5, pp. 2493–2518, 2011. https://doi.org/10.1214/11-AOAS493

[16] F. R. Halim and M. A. R. Novita Afdal, “Sentiment Analysis to Assess Customer Retention on Instagram Social Media,” in *4th Int. Conf. on Emerging Smart Technologies and Applications (eSmarTA)*, 2024.

[17] F. Prastiani, H. Fakhrurroja, and F. Hamami, “Social Media Sentiment Analysis for Local Water Company Customers Using a Support Vector Machine Algorithm,” in *10th Int. Conf. on ICT for Smart Society (ICISS)*, 2023.

[18] M. H. Khasawneh, A. Shuhaiber, and M. Al Khasawneh, “A Comprehensive Model Of Factors Influencing Consumer Attitude Towards And Acceptance Of SMS Advertising,” 2013.

[19] N. F. Alshammari and A. A. AlMansour, “Aspect-based Sentiment Analysis for Arabic Content in Social Media,” in *Int. Conf. on Electrical, Communication, and Computer Engineering (ICECCE)*, Istanbul, 2020.

[20] A. Mountassir, H. Benbrahim, and I. Berrada, “Some Methods to Address the Problem of Unbalanced Sentiment Classification in an Arabic Context,” in *2012 Colloquium in Information Science and Technology*, 2012.

[21] A. Al-Sallab *et al.*, “AROMA: A recursive deep learning model for opinion mining in Arabic as a low resource language,” 2017.

[22] A. A. Hnaif, E. Kanan, and T. Kanan, “Sentiment analysis for Arabic social media news polarity,” *Intell. Autom. Soft Comput.*, vol. 28, pp. 107–119, 2021. https://doi.org/10.32604/iasc.2021.015939

[23] A. M. Azmi and S. M. Alzanin, “Aara’ - A system for mining the polarity of Saudi public opinion through e-newspaper comments,” *J. Inf. Sci.*, vol. 40, pp. 398–410, 2014. https://doi.org/10.1177/0165551514524675

[24] S. Alowaidi, M. Saleh, and O. Abulnaja, “Semantic Sentiment Analysis of Arabic Texts,” 2017.

[25] S. Khalid, S. Wu, and F. Zhang, “A multi-objective approach to determining the usefulness of papers in academic search,” *Data Technol. Appl.*, vol. 55, no. 5, pp. 734–748, 2021.

[26] S. Khalid and S. Wu, “Supporting scholarly search by query expansion and citation analysis,” *Eng. Technol. Appl. Sci. Res.*, vol. 10, no. 4, pp. 6102–6108, 2020.

[27] S. Khalid, S. Khusro, and I. Ullah, “Crawling ajax-based web applications: Evolution and state-of-the-art,” *Malays. J. Comput. Sci.*, vol. 31, no. 1, pp. 35–47, 2018.

[28] S. Bano and S. Khalid, “BERT-based extractive text summarization of scholarly articles: A novel architecture,” in *2022 Int. Conf. on Artif. Intell. of Things (ICAIoT)*, pp. 1–5, IEEE, 2022.

[29] S. Bano *et al.*, “Summarization of scholarly articles using BERT and BiGRU: Deep learning-based extractive approach,” *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 35, no. 9, p. 101739, 2023.

[30] S. Alaparthi and M. Mishra, “Bidirectional Encoder Representations from Transformers (BERT): A sentiment analysis odyssey,” *J. Mark. Anal.*, 2021. https://doi.org/10.1057/s41270-021-00181-5

[31] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in *Proc. 2019 Conf. of the NAACL: Human Lang. Technol.*, 2018.

[32] H. Alamoudi *et al.*, “Arabic Sentiment Analysis for Student Evaluation using Machine Learning and the AraBERT Transformer,” *Eng. Technol. Appl. Sci. Res.*, vol. 13, no. 5, pp. 11945–11952, 2023. https://doi.org/10.48084/etasr.6347

[33] H. Wang, L. Wang, and C. Liu, “Employee competitive attitude and competitive behavior promote job-crafting and performance: A two-component dynamic model,” *Front. Psychol.*, vol. 9, 2018. https://doi.org/10.3389/fpsyg.2018.02223

[34] O. Adel, K. M. Fathalla, and A. A. ElFarag, “MM-EMOR: Multi-Modal Emotion Recognition of Social Media Using Concatenated Deep Learning Networks,” *Big Data Cogn. Comput.*, vol. 7, 2023. https://doi.org/10.3390/bdcc7040164

[35] B. Khasoggi, Ermatita, and Samsuryadi, “Efficient mobilenet architecture as image recognition on mobile and embedded devices,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 16, pp. 389–394, 2019. https://doi.org/10.11591/ijeecs.v16.i1.pp389-394

[36] J. Bethge *et al.*, “MeliusNet: Can Binary Neural Networks Achieve MobileNet-level Accuracy?” *arXiv preprint arXiv:2001.05936*, 2020.

[37] M. S. Al Reshan *et al.*, “Detection of Pneumonia from Chest X-ray Images Utilizing MobileNet Model,” *Healthcare (Switzerland)*, vol. 11, 2023. https://doi.org/10.3390/healthcare11111561