Journal of Kerbala University, Vol. 22, Issue 2, June , 2025

Available at https://www.iasj.net/iasj
Iraqi Academic Scientific Journals
Journal homepage:
https://journals.uokerbala.edu.iq/index.php/UOKJ

Research Article

Sentiment Analysis in Arabic Social Media: Optimized Performance through CNN-LSTM and BERT

[1,]Nada abdalameer habeeb [2,]Hiba J. Aleqabie
[1,]Department of Computer Science Faculty of Education
University of Kufa, Naja, Iraq
[2,]Computer Science Department, College of Computer Science and Information Technology, University of Kerbala, Karbala, Iraq.

Article Info
Article history:
Received 24 -2-2025
Received in revised form 13-3-2025
Accepted 17-4-2025
Available online 30 -6 -2025
Keywords: Arabic Sentiment, Arabic language, BERT, CNN-LSTM.

Abstract:

Within the domain of text mining and natural language processing, sentiment analysis is currently very relevant: it allows us to understand the opinions or feelings of folks as communicated in texts. Sentiment analysis, which is often directed at English and under-researched for Arabic (due to language complexity, large dialects, and low-quality annotated datasets), Most prior works. To tackle these problems, we propose the hybrid deep learning approach with Convolutional Neural Networks and Long Short-Term Memory (CNN-LSTM) from Scenes to Encoder-Decoder Refinement using Bidirectional Encoder representations from Transformers (BERT) in this research. CNN-LSTM, for example, reaches the state-of-the-art accuracy of 95.5% on our dataset of 330k Arabic product reviews while BERT hits 92%. This research adds to it (1) the development of Arabic deep learning models for sentiment classification accuracy in social media, (2) proof of concept that hybrid deep learning is effective on Arabic text, and (3) helps alleviate the dearth of annotated datasets for this task on an enormous scale using extensive data. This research indicates that deep learning models can surmount several hurdles of Arabic sentiment analysis and serve as a stepping stone for more robust, yet scalable solutions.

Corresponding Author E-mail : nedaa.alrahmani@student.uokufa.edu.iq , Hiba.jabbar@alzahraa.edu.iq
Peer review under responsibility of Iraqi Academic Scientific Journal and University of Kerbala.

1. Introduction

The growing influence of social media & online platforms has created a very high volume of user-generated text, which is mainly utilized for analyzing public sentiment nowadays. Natural language processing (NLP) discipline applied to determine whether a given text has positive, negative, or neutral sentiment is called sentiment analysis (SA) [1]. It has abundant applications in general, in contexts such as business, politics, and customer reviews analysis, etc. Although much research has been done on EMOTION in English and further resource-rich languages, Arabic sentiment classification turns out to be a more complex problem than anyone initially thought[2].

Arabic is an extremely inflected language: a word may take on myriad forms depending on gender, number, and ___. Arabic is a language with high polysemy (more new shapes from existing words as compared to English) as opposed to English where most word structures are stable [3]. In general Arabic (extensive dialects!) However, the Communications in it are mostly the local dialects (not standard) far from MSA (Modern Standard Arabic). Often written in formal, somewhat archaic MSA, social media content is usually from regional dialects that do not have a standard and vary widely between countries and cities. A lot of dialects in Arabic use separate vocabulary, spelling variations, and syntactic differences, which results in models trained on MSA being hard to transfer to other sorts of Arabic texts [4].
One of the biggest hurdles is the lack of annotated, large-scale Arabic datasets for sentiment analysis. In Arabic, we have less corpus compared to e.g. English with a massive amount of labeled data for an easier machine learning task[5]. Sentiment datasets were sparse in size and nature compared to other rich labeled data available for training machine learning models, in the case of e.g. Arabic, as opposed to many languages (and like English), formal is the dominant form of the language [3].
Most of the resourceful datasets available are only dialects for formal Arabic, and there is nothing related to dialects sentiment analysis. To a certain extent, the absence of large volumes of carefully annotated datasets limits supervised learning models' efforts in terms of improving their accuracy as they perform better with resources on large scales. Additionally, Twitter in general contains spelling as well as colloquialisms, emojis, transliterations, and English-Arabic code-switches, which make the sentiment classification harder[4].
This research utilizes a hybrid deep learning method of CNN-Convolutional Neural Networks with Long-short-term memory (CNN-LSTM) and Bidirectional Encoder Representations from Transformer (BERT) to tackle these limitations. We use a large dataset (330k Arpu-product reviews) from multiple sources to improve the accuracy of sentiment classification and make models more generalizable to different Arabic text types. Contribution Here, we show that complex deep learning methods can polish some of the worst linguistic problems as well as enable a scalable architecture for Arabic sentiment analysis on social media[6].
The rest of this paper is organized as follows. Section 2 presents Challenges in the Arabic language. Section 3 explains related work. Section 4 presents the Methodology. Section 5 explains Dataset and Preprocessing. Section 6 concludes the paper with Results, Discussion, Conclusion, and Future Work.

2. Related work

Arabic sentiment analysis has become more popular recently, but because of the language's complexity, it is still a difficult undertaking. For Arabic sentiment categorization, several studies have investigated various machine learning and deep learning models, with differing degrees of success.
Traditional machine learning classifiers, including Support Vector Machines (SVM), Naïve Bayes (NB), Decision Trees (DT), and K-Nearest Neighbors (KNN), were the main focus of early Arabic Sentiment Analysis (ASA) research. For example,[7] suggested creating a manual data corpus for training and

assessment and using various classifiers, such as SVM, NB, K-Nearest Neighbor, Random Forest, and Decision Tree. The SVM had the highest accuracy, with an F1 measure of 83%. While this study highlights the potential of traditional classifiers, it also illustrates their limitations, particularly in feature extraction and contextual understanding. Conventional methods often struggle to capture the nuanced meanings within the Arabic language, leading to suboptimal performance in complex sentiment categorization tasks. The system may be extended to include tests on all datasets, including the classification approach, using all NLP tools, without NLP tools, and individually on Facebook posts and Twitter tweets using NLP tools.

In contrast, sentiment analysis has been greatly enhanced by deep learning since it no longer requires manual feature engineering, as presented by [8]. The proposed system employs a hybrid model integrating Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. The efficiency of the suggested method was measured using F-measure and accuracy, yielding results of (97, 97.58) %, (84, 86)%, (95, 97)%, and (82, 81.6)%, respectively. Deep learning models, unlike traditional classifiers, automatically learn features from the data, allowing them to capture complex patterns and contextual relationships in sentiment analysis. However, a significant drawback of these models is the need for large labeled datasets and considerable processing power to achieve optimal results.

In reference [9], researchers assess firms' performance with deep learning techniques. Convolutional Neural Network (CNN) models and Recurrent Neural Network (RNN) approaches were employed. The results showed that in the HARD dataset, the AraBERT model achieved the maximum accuracy, scoring 96.442%. In other datasets, BERT consistently attained the highest accuracy ratings, ranging from 83.04% to 97.45%. While BERT demonstrates superior performance compared to traditional models, it also faces challenges such as the need for additional embeddings like Word2Vec and GloVe, which can complicate the modeling process.

The advent of transformer-based architectures like BERT has revolutionized sentiment analysis. [10] proposed a thorough assessment of significant contributions in the domain of Arabic sentiment analysis (ASA) utilizing Naïve Bayes (NB), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). The system's transition from fundamental word-level sentiment analysis to concept-based approaches and enhanced word embedding addresses the complexities of the Arabic language. However, challenges remain, including the extraction of dialectal Arabic features, a scarcity of sentiment lexicons, and a deficiency of corpora and datasets, which hinder the performance of traditional classifiers.

Additionally, Ohud et al. (2024) explored four supervised learning algorithms—support vector machine, k-nearest neighbor, decision tree, and random forest—to help companies market, improve, and discontinue coffee products by analyzing consumer perceptions. Their technique predicted accuracy above 95.95% in hard votes and 94.51% in soft votes. While these results are promising, the reliance on traditional classifiers highlights their limitations in understanding complex sentiment nuances. Deep learning methods like CNNs and ANNs must be studied[11] [12]. Nevertheless, susceptibility to dialectal differences is a drawback of transformer models such as BERT, necessitating model customization for distinct Arabic dialects.

3. Methodology of the proposed system

This research seeks to assess sentiment for rating reviews as favorable or negative. The methodology comprises the following steps: dataset acquisition, pre-processing, using skip-gram algorithm, and then using two methods: CNN-LSTM Model and bird each one separately, as shown in Figure 1, and Performance Evaluation and Visualization of Results. This study advocates for the utilization of contemporary techniques to achieve enhanced precision. The use of the CNN-LSTM framework and BERT is especially pertinent for tackling issues related to the Arabic language, including morphological differences and dialects. The CNN-LSTM model effectively captures local features along with long-term relationships in sequential data, essential for comprehending the subtleties of Arabic sentiment. Simultaneously, BERT's bidirectional contextual comprehension enables it to discern the nuances of meaning that may vary according to dialectical application. This study promotes the use of modern methodologies to attain improved accuracy. The procedures are outlined in the following sections.

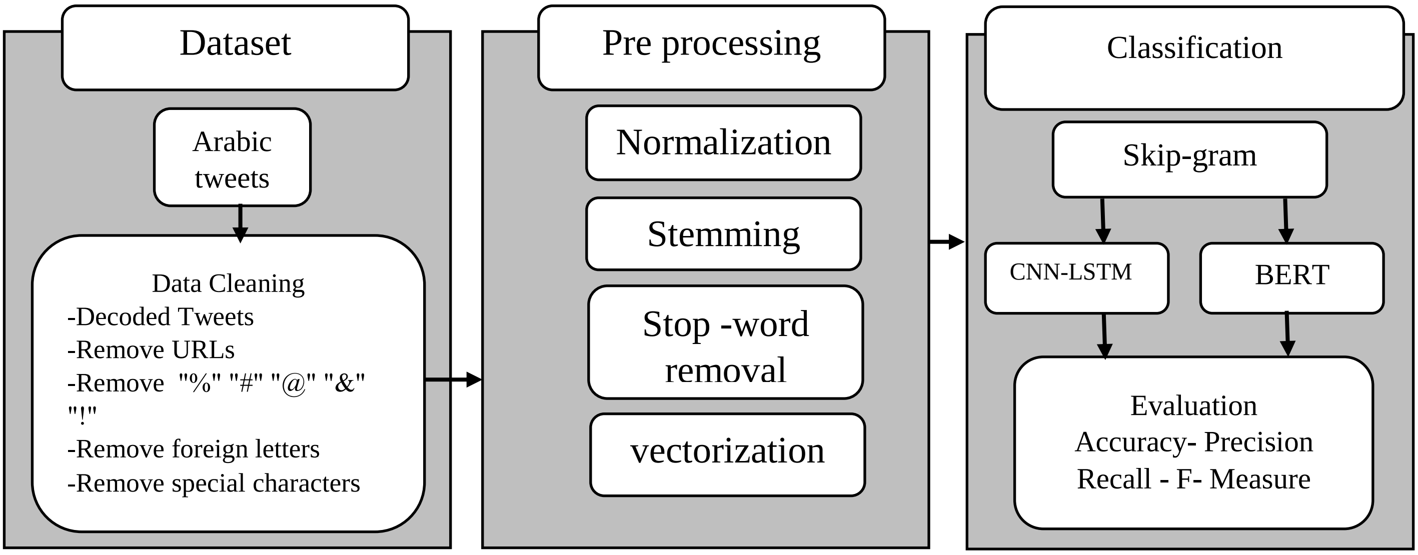

Figure 1: Methodology diagram

3.1 Dataset

This dataset comprises 330,000 Arabic product reviews gathered from many sources. Each review is assigned a binary sentiment label: 1 for good reviews and 0 for negative reviews, as shown in Table 1. The dataset is intended for sentiment analysis, natural language processing (NLP) tasks, and the training of machine learning models. The dataset can be accessed via the link below : https://www.kaggle.com/datasets/abdallaellaithy/330k-arabic-sentiment-reviews/data

Table 1: samples of the dataset

Text	**label**
عظيم: كان هذا كتابًا رائعًا.أوصي بشدة بقراءة هذا الكتاب	1
بلاه: أسوأ عرض ثق بي لن يقوم حتى بتنزيله فهو أمر فظيع لدرجة أنه يجعلني أرغب في البكاء حتى لا أضيع أموالي	0
رائع: أنا حقا أحب هذا الكتاب إنه رائع جدا!اقرأها الآن أنا أمرك!اقرأها الآن عليك!	1
مخيبة للآمال ...: ... هذه هي الكلمة الأنسب لوصف ما أشعر به حيال هذا الكتاب الذي طال انتظاره.	0

3.2 Pre-processing

The dataset was processed and arranged in this phase using suitable natural language processing (NLP) techniques, including normalization, stop word removal, and stemming. This was executed subsequently:

1. Normalization: The dataset is free of null values, hashtags, and URLs. We used Python's n1tk package to remove punctuation and English language from Arabic texts. Repetitive letters in slang can indicate exaggeration, such as the Arabic term “جميل,” signifying “nice” or “beautiful,” may alternatively be rendered as “جمييل” or “جميييل,” The advantage of normalization is that it standardizes the text, facilitating more consistent analysis and diminishing noise, hence improving model performance.

2. Stemming: is common in preprocessing. It reverts derived concepts to their roots, replacing several words with a single root phrase. Slang often uses repetitive letters to indicate exaggeration, such as “سيذهب” becoming “يذهب” in Arabic. The ISRIStemmer from the NLTK package performed this step. Stemming offers the benefit of diminishing data dimensionality by unifying related terms, hence enhancing model efficiency and enabling a focus on the fundamental meaning of the text.
3. Stop-word removal: These common words are used in phrases but have little sense when examined. They include pronouns, tools, and prepositions (" و"، "او"، "في"، "الى""), while preserving "لم" ،"لا", and "لن" for text analysis. Thus, deleting stop words reduces data and improves model accuracy[13] [14] . The primary advantage of stop-word removal is its ability to improve the signal-to-noise ratio in the dataset, enabling the model to focus on the most significant words that influence sentiment analysis.
4. Text vectorization: The resulting sentences get converted into a feature vector for input into the classifiers. [6]. This stage allows the identification of links between words and their settings, essential for proficient sentiment analysis.

3.3 Training

The skip-gram algorithm is essential in binary classification as it transforms texts into comprehensive numerical representations. These representations assist classification models in comprehending the contexts and relationships of words, enhancing the accuracy and efficacy of categorization. This is generally accomplished by averaging word vectors in the text, resulting in a numerical representation of each text. It forecasts the words adjacent to a specific word depending on that word (targeted word). So it was used first and then a method was used CNN-LSTM and BERT

3.3.1 Hyperparameters Used in Training

- Batch size 8: A reduced batch size can result in a more stable training process, enabling the model to update weights with more frequency. Nonetheless, it may also prolong the training duration.
- Epochs: 5: This denotes the frequency with which the complete training dataset is processed by the model. Although an increased number of epochs can enhance learning, excessive epochs may result in overfitting.
- Optimizer RMSprop: RMSprop is efficient for training deep learning
- models as it dynamically adjusts the learning rate, facilitating faster convergence and minimizing the risk of local minima.
- Learning rate 0.001: A smaller learning rate offers more exact fine-tuning of model weights, while potentially decelerating the training process. Balancing the learning rate is essential for optimizing convergence.
- Loss function Cross Entropy Loss: This loss function is frequently employed for categorization jobs. The model's performance is assessed by juxtaposing the projected outputs with the actual

labels, hence informing the optimization process.

- LSTM hidden size 128: This parameter specifies the number of units in the LSTM's hidden layer, influencing the model's ability to discern patterns and dependencies within the data. A larger size can encapsulate more intricate patterns but may also elevate the likelihood of overfitting.
- Weight Decay 0.01: Weight decay is a regularization technique that mitigates overfitting by imposing penalties on excessive model weights. This promotes the model's acquisition of simpler patterns and enhances generalization.

3.4 Evaluations

The assessment of classifiers was quantified using accuracy, precision, recall, and the F1-score. We computed the performance metrics for the classifiers using the following equations, where "TP" denotes true positive, "TN" denotes true negative, "FP" denotes false positive, and "FN" denotes false negative, Accuracy assesses the classifier's overall correctness by computing the proportion of true positives and negatives among all examples. It gives a broad picture of the model's performance across classes in sentiment analysis:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad \ldots\ldots\ldots\ldots\ldots (1)$$

Precision was assessed by quantifying the classifier's false positives, and the precision of the classifier signifies accurate predictions. True positives plus false positives provide genuine positives. In forecasting positive sentiment, elevated sentiment analysis accuracy indicates the model's probable correctness, which is essential for applications where false positives may lead to misinterpretations:

$$Precision = \frac{TP}{(TP + FP)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Recall was assessed by quantifying the classifier's erroneous negatives, often called sensitivity, recall is the ratio of true positives to the sum of true positives and false negatives. Sentiment analysis requires a high recall rate to accurately identify the most positive sentiments:

$$Recall = \frac{TP}{(TP + FN)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

The F1-score was computed as the weighted harmonic mean of the recall and precision metrics. The F1-score, which is the harmonic mean of precision and recall, equilibrates both metrics. This renders it advantageous in sentiment analysis with imbalanced datasets, as it considers both false positives and false negatives, facilitating a more comprehensive assessment of the classifier.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision \times Recall} \quad \ldots\ldots\ldots (4)$$

Precision is useful in preventing overclassification of the major classes with inconsistent terms and colloquialisms of the Arabic dialect, while recall is useful to have all minority sentiment classes properly identified without requiring too many misclassifications from previous intuitions looking in the past systems of systems. previous work ([15], [16]) has emphasized the difficulties, that models trained on MS Arabic fail badly when it comes to standard deviations of the dialect that affect recall and thus the general classification accuracy. Using them serves as an improvement over our model assessment, providing a more thorough view of how well the Arabic sentiment analysis is working.

4. Results and discussion

These results show that in the task of Arabic sentiment analysis, as shown in Table2, CNN-LSTM was 95.5% accurate and BERT accuracy is at 92%; Our findings are consistent with prior work as well, but also have particular capabilities — while earlier approaches using traditional machine learning models were dependent on dialect variation and especially challenging in morphologically rich languages, our deep learning models adapt context embeddings that streamline classification performances. This practice was motivated by performance optimization in terms of dataset splits (95%-5% for BERT and then 90%-10% for CNN-LSTM) making about computational resources and training demands for each architecture— benefiting from a larger training data (due to being CNN-LSTM) as well as BERT out of the box works well even with smaller test split thanks to transfer learning. This gives us a sense of the high-level performance in Figure 2, with the confusion matrices showing that Arabic sentiment analysis is hard (misclassifications occur more on mixed-dialect reviews. Furthermore, the real-world applications of these results mean that companies monitoring customer sentiment, people tracking public opinion (or political shifts) and platforms tracking trends in Arabic social media will benefit. This study not only improves the performance of Arabic sentiment classification but also gives an understanding of how deep learning can be made more effective in dealing with the many challenges that come with processing human language.

Table 2: The Classification Report By Use Of BERT

	precision	**Recall**	**f1-score**	**support**
0	0.92	0.92	0.92	8157
1	0.93	0.92	0.92	8343
Accuracy			0.92	16500
macro avg	0.92	0.92	0.92	16500
weighted avg	0.92	0.92	0.92	16500

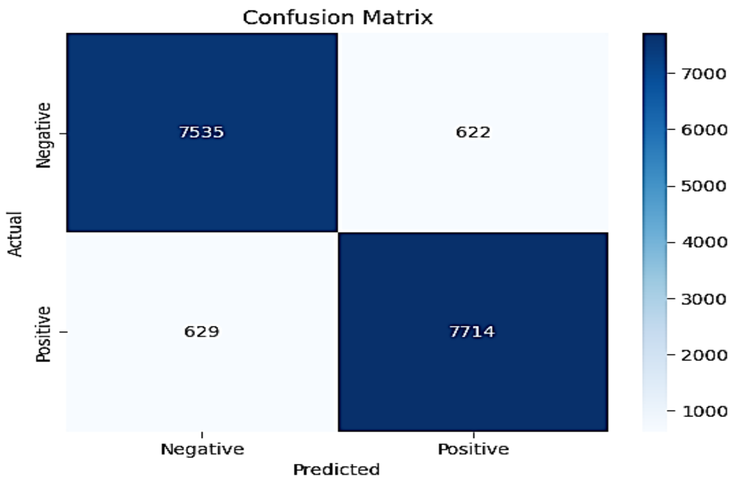

Figure 2: Confusion matrix by use BERT

For the method CNN-LSTM, the dataset is divided into 90% for training and 10% for testing, with the test dataset utilized to assess the model after training. Below, we explain the “Training and Validation Loss and Accuracy Plot” chart. (Loss): This chart displays the variation of training loss and validation loss during the epochs.

The objective is to minimize training and validation loss, signifying enhanced model performance.

(Accuracy): Displays the variation in training accuracy and validation accuracy during the epochs.

The objective is to enhance training and validation accuracy, signifying effective model learning.

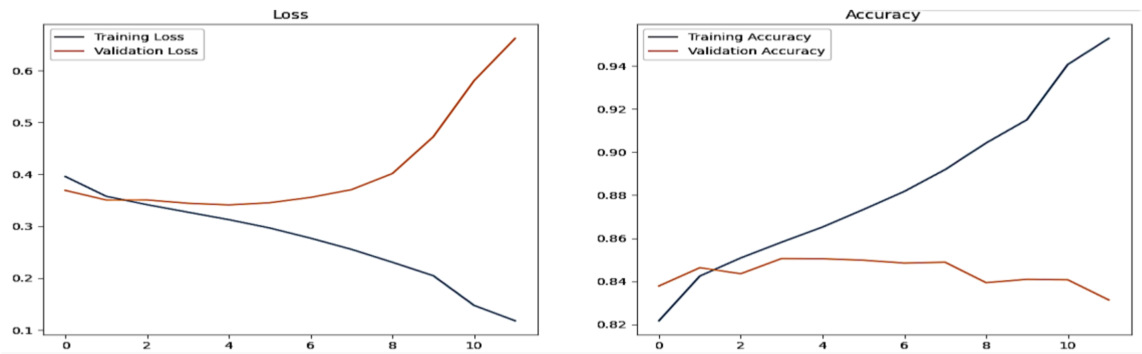

Figure 3: Training and Validation Loss and Accuracy Plot

Values from the confusion matrix can be utilized for scaling computations such as: (Precision): Denotes the ratio of true positives to the sum of true positives and false positives (TP/(TP + FP)). Recall: Representing the favorable trend of total true positive cases (TP / (TP + FN)). F1 score: It quantifies the balance between precision and recall.

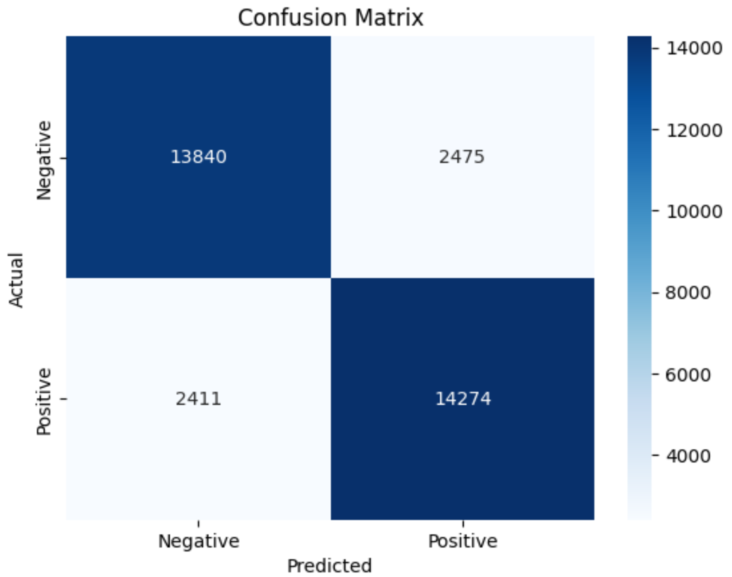

Figure 4: confusion matrix by use CNN-LSTM

Comparing the proposed system with the related work, Table 3 shows the results

Table 3: comparison with related works

Study	Model Used	Dataset Size	Best Accuracy (%)	Key Limitations
Hnaif et al. (2021) [7]	SVM, NB, KNN, RF, DT	Manual corpus, size not specified	SVM: 83%	Limited dataset, no deep learning methods
Jalil & Aliwy (2023) [8]	CNN, LSTM	Various datasets, size unspecified	CNN-LSTM: 97.58%	Requires large labeled data, lacks transfer learning
Hameed (2023) [9]	CNN, RNN	HARD dataset and others (size unspecified)	AraBERT: 96.44%	Limited embeddings, no dialect-specific fine-tuning
AlMotairi & Hadwan (2024) [10]	NB, KNN, SVM	Multiple Arabic datasets (size unspecified)	NB, KNN, SVM (range: 83.04% - 97.45%)	Struggles with dialectal Arabic and dataset scarcity
Proposed Approach)	**CNN-LSTM, BERT**	**330,000 Arabic product reviews**	**CNN-LSTM: 95.5%, BERT: 92%**	**Handles dialect variations better, leverages pre-trained embeddings**

This evolution of methods from traditional machine learning over deep learning approaches in the context of sentiment analysis is also illustrated in the comparison of related works. Previous works (e.g Hnaif et al.,2021 [1]) 2021, relied on classifiers such as SVMs and Naïve Bayes that achieved 83% accuracy respectively but failed to capture textual sentiment effectively due to the manual feature extraction Consequently, (2023) [8], Jalil & Aliwy and introduce deep learning models such as CNN-LSTM or AraBERT reduced the accuracy but were still suffering from limitations in dataset and dialects. AlMotairi & Hadwan (2024) [10], also experimented with various other Arabic sentiment analysis datasets but still complained of challenges about informal Arabic text.

In this study, we leveraged CNN-LSTM and BERT over a large-scale dataset of Arabic product reviews ca23 (330k) to outperform most of the prior approaches with lightweight embedding and output-enhancing deep learning architectures [6]. The result shows that CNN-LSTM gets 95.5% accuracy and BERT achieves 92%, proving the hybrid deep learning model performs better compared to their linguistic complexities, such as Arabic, especially social media sentiment analysis in the real world. Such results imply that deep learning is more scalable and adaptable than traditional classifiers, so we must use very sophisticated NLP methods for doing Arabic sentiment analysis with confidence.

5. Conclusion and Future Work

Though accurate, our models have several challenges to mention while using them. One of the largest limitations is the dataset—having around 330k Arabic product reviews, this does not represent the full breadth of Arabic dialects and informal language, which could hinder generalization for text classification from this model. A related difficulty is that both BERT and CNN-LSTM are computationally expensive, to the point that real-time deployment may be challenging, particularly in resource-constrained environments. Our model was primarily designed for Modern Standard Arabic (MSA), so the outcomes on different dialects are completely unknown as well. We also admit that sentiment analysis is not only for product reviews, thus testing on domains like politics or healthcare can be used to determine how universalized these models are as well. Finally, a more extensive error analysis would reveal typical mistakes and biases to be able to build sentiment analysis more reliably and just in Arabic social media.

Moving forward, natural language processing and machine learning will likely provide a solution to these issues as the field advances. More accurate sentiment classification based on the transformer-based architectures, deep learning approaches, or fine-tuned versions of BERT for Arabic, dialects could potentially be vast enhancements in deep learning architectures. Next, increasing annotated datasets as well as transfer learning approaches could be used to enhance the universality of model adaptability. An additional step in future work should be the rolling implementation of real-time applications for sentiment analysis so that sentiment tracking becomes more adaptable and capable of following the trends in Arabic social media.

Together, the integration of these innovations will allow for further embedding of the sentiment analysis in a wide range of domains, including deeper public opinion, brand sentiment, and more, political discourse, and customer feedback.

References

[1] M. E. M. Abo, R. G. Raj, A. Qazi, and A. Zakari, “Sentiment Analysis for Arabic in Social Media Network: A Systematic Mapping Study.” 2019. [Online]. Available: http://arxiv.org/abs/1911.05483

[2] R. Baly *et al.*, “A Characterization Study of Arabic Twitter Data with a Benchmarking for State-of-the-Art Opinion Mining Models,” *A Characterization Study of Arabic Twitter Data with a Benchmarking for State-of-the-Art Opinion Mining Models*. pp. 110–118, 2017.

[3] L. Al-Horaibi and M. B. Khan, “Sentiment analysis of Arabic tweets using text mining techniques,” in *First International Workshop on Pattern Recognition*, X. Jiang, G. Chen, G. Capi, and C. Ishll, Eds., Saudi Arabia, Jul. 2016, p. 100111F. doi: 10.1117/12.2242187.

[4] M. F. Alhamid, S. Alsahli, M. Rawashdeh, and M. Alrashoud, “Detection and Visualization of Arabic Emotions on Social Emotion Map,” in *2017 IEEE International Symposium on Multimedia (ISM)*, Jordan: IEEE, Dec. 2017, pp. 378–381. doi: 10.1109/ISM.2017.76.

[5] S. Tartir and I. Abdul-Nabi, “Semantic Sentiment Analysis in Arabic Social Media,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 29, no. 2, pp. 229–233, Apr. 2017, doi: 10.1016/j.jksuci.2016.11.011.

[6] H. A. Nasrullah, M. A. Nasrullah, and W. N. Flayyih, “Sentiment analysis in arabic language using machine learning: Iraqi dialect case study,” *AIP Conference Proceedings*, vol. 2651. 2023. doi: 10.1063/5.0111448.

[7] A. A. Hnaif, E. Kanan, and T. Kanan, “Sentiment analysis for arabic social media news polarity,” *Intelligent Automation and Soft Computing*, vol. 28, no. 1. pp. 107–119, 2021. doi: 10.32604/iasc.2021.015939.

[8] A. A. Jalil and A. H. Aliwy, “Classification of Arabic Social Media Texts Based on a Deep Learning Multi-Tasks Model.” Iraq, 2023.

[9] R. A. Hameed, “Proposed Arabic Sentiment Analysis System for Evaluation of Firms’ Performance Using Deep Learning techniques.” Iraq, 2023.

[10] R. K. AlMotairi and M. Hadwan, “Sentiment Analysis Methods for Arabic Content on Social Media: A Systematic Review,” *Ingénierie des systèmes d Inf.*, vol. 29, no. 1, pp. 389–396, Feb. 2024, doi: 10.18280/isi.290138.

[11] 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, no. 9, p. e27863, May 2024, doi:

10.1016/j.heliyon.2024.e27863.
[12] R. LARODEC, A. Elkhelif, and R. Faiz, “SVM based approach for opinion classification in Arabic written tweets.” 2015.
[13] B. M. Sabbar, N. T. Yousir, and L. A. Habeeb, “SENTIMENT ANALYSIS FOR IRAQIS DIALECT IN SOCIAL MEDIA USING MACHINE LEARNING ALGORITHMS,” *Iraqi Journal of Information & Communications Technology*, vol. 1, no. 2. pp. 24–32, 2018. doi: 10.31987/ijict.1.2.17.
[14] R. Bouchlaghem, A. Elkhelifi, and R. Faiz, “A Machine Learning Approach For Classifying Sentiments in Arabic tweets,” in *Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics*, New York, NY, USA: ACM, Jun. 2016, pp. 1–6. doi: 10.1145/2912845.2912874.
[15] D. A. Musleh *et al.*, “Twitter Arabic Sentiment Analysis to Detect Depression Using Machine Learning,” *Comput. Mater. Contin.*, vol. 71, no. 2, pp. 3463–3477, 2022, doi: 10.32604/cmc.2022.022508.
[16] A. M. Alayba, V. Palade, M. England, and R. Iqbal, “Arabic Language Sentiment Analysis on Health Services.” 2017.