

## Evaluation of Hotel Performance with Sentiment Analysis by Deep Learning Techniques

<https://doi.org/10.3991/ijim.v17i09.38755>

Rafeef A. Hameed<sup>1</sup>(✉), Wael J. Abed<sup>2</sup>, Ahmed T. Sadiq<sup>1</sup>

<sup>1</sup> Computer Sciences Department, University of Technology, Baghdad, Iraq

<sup>2</sup> Computer Techniques Engineering Department, Al-Mustaqbal University College, Hillah, Iraq

cs.21.23@grad.uotechnology.edu.iq

**Abstract**—The subject of sentiment analysis through social media sites has witnessed significant development due to the increasing reliance of people on social media in advertising and marketing, especially after the Corona pandemic. There is no doubt that the prevalence of the Arabic language makes it considered one of the most important languages all over the world. Through human comments, it can know things if they are positive or negative. But in fact, the comments are many, and it takes work to evaluate the place or the product through a detailed reading of each comment. Therefore, this study applied deep learning approaches to this issue to provide final results that could be utilized to differentiate between the comments in the dataset. Arabic Sentiment Analysis was used and gave a percentage for each positive and negative commentary. This work used eight methods of deep learning techniques after using Fast Text as embedding, except Ara BERT. These techniques are the transformer (AraBERT), RNN (Long short-term memory (LSTM), Bidirectional long-short term memory (BI-LSTM), Gated recurrent units (GRUs), Bidirectional Gated recurrent units (BI-GRU)), CNN (like ALEXNET, proposed CNN), and ensemble model (CNN with BI-GRU). The Hotel Arabic Reviews Dataset was utilized to test the models. This paper obtained the following results. In the Ara BERT model, the accuracy is 96.442%. In CNN, like the Alex Net model, the accuracy is 93.78%. In the suggested CNN model, the accuracy is 94.43%. In the suggested LSTM model, the accuracy is 95%. In the suggested BI-LSTM model, the accuracy is 95.11%. The accuracy of the suggested GRU model is 95.07%. The accuracy of the suggested BI-GRU model is 95.02%. The accuracy is 94.52% in the Ensemble CNN with BI-GRU model that has been proposed. Consequently, the AraBERT outperformed the other approaches in terms of accuracy. Because the AraBERT has already been trained on some Arabic Wikipedia entries. The LSTM, BI-LSTM, GRU, and BI-GRU, on the other hand, had comparable outcomes.

**Keywords**—Arabic sentiment analysis, NLP, deep learning, embedding, CNN, RNN, AraBERT

## 1 Introduction

Social media today offers a fantastic platform for expressing thoughts and exchanging firsthand knowledge about various occasions, goods, and services. For internet users to choose the best service or product to buy, such helpful information sources are highly interesting. Indeed, opinions are significant because they are impartial, independent, and founded on accurate user experiences with a particular good or service. The feedback from users is also valuable to businesses since it allows them to gauge client happiness and enhance the caliber of their goods and services. As a result, it is simple to gather data and distinguish between positive and negative emotions, which is a hotly debated study topic. Comparatively, few studies have been done on sentiment analysis of Arabic literature because Arabic is difficult to learn. This is because the majority of the literature on sentiment analysis focuses on the English language. Arabic Sentiment Analysis (ASA) has been an important research subject due to the recent enormous more challenging [1]. There is little study on Arabic-related feelings, attitudes, emotions, and ideas [2]. The main goal of the ASA assignment is to assign Arabic text to predetermined classes based on its content. Text representation is an important process that impacts ASA performance, and as contextual embedding models can account for both context and word meaning, they are useful for learning universal sentence representations. Often known as opinion mining, SA is the process of determining if a writer has a negative or positive attitude about a certain thing [3]. The main contribution of this paper is present a proposed CNN model for ASA. In addition, this paper uses other deep learning models for ASA that are based on AraBERT, CNN, LSTM, BiLSTM, and GRU, BiGRU. Fast Text [4, 5] embedding has been used for text representation. Datasets from various sources were used to train the deep-learning models. The following components comprise this article: Section two, which contains the relevant work; Section three, which gives the recommended technique; Section four, which has the experimental findings; and Section five, which contains the conclusion.

## 2 Related work

In [6], a one-layer CNN architecture, two LSTM layers, and a deep learning model for Arabic sentiment analysis were expertly coupled. Fast Text word embedding is used to support the input layer of this design. The investigations on a multi-domain corpus revealed that this model performed very well in terms of precision, recall, F1-Score, and accuracy, scoring 89.10%, 92.14%, 92.44%, and 90.75%, respectively. The impact of word embedding techniques on Arabic sentiment categorization was carefully examined in this study, and it was found that the Fast Text model is a better choice for learning semantic and syntactic information. NB and KNN classifiers are used to evaluate the effectiveness of the proposed model. The results showed that SVM is the best-performing classifier, with an accuracy improvement of up to + 3.92%. Because of the effectiveness of the CNN in features extraction and the recurrent nature of LSTM. In [7], the author employed a long short-term memory recurrent neural network (LSTM), a convolutional neural network (CNN), and an ensemble model incorporating both

models to extract semantic information for short Arabic text at the word and character levels. A dataset comprised of dialectal Arabic corpora and Modern Standard Arabic corpora that were gathered from Twitter was used to train and test the models. The values obtained ranged from 89.7% to 69.7%. The ensemble model had the test set's highest accuracy rating of 96.7%.

In [8], Compared and evaluated various sentiment analysis models on Arabic tweets in the article. Performance of four deep learning models—CNN, LSTM, BI-LSTM, GRU, and a hybrid model (BI-LSTM + GRU) with three text representation techniques—was empirically evaluated (i.e., AraVec, FastText, AraBERT). The proposed model (BI-LSTM + GRU) using the AraBERT model has the best accuracy of these models, coming in at 0,9434. The examination of the deep learning model outputs demonstrates unmistakably that for our dataset, the hybrid network performs better than other models for various word embeddings, and their accuracy is higher than that of other models. In [9], offered a tagged corpus of 40k Arabic tweets on a variety of subjects, such as politics, sports, health, sarcasm, proverbs, and poetry. The article also used three deep-learning methods for the suggested corpus. In particular, the paper tested how well the corpus performed using CNN, LSTM, and RCNN. The LSTM outperformed CNN with an accuracy of 75.72% and RCNN with an accuracy of 78.46% using the word embedding approach as the input layer to the three models. With an accuracy of 88.05% after using a data augmentation strategy on the corpus, LSTM has demonstrated a further improvement.

In [10], They solved the issue with Arabic Text Sentiment Analysis. This study takes advantage of a deep learning model's performance-improving effects on the Arabic Sentiment Analysis system. To forecast the sentiment of the Arabic text, they employed the BI-LSTM deep learning model, which has the capacity to extract contextual information. On six benchmark datasets, experiments are run to gauge how well the proposed methodology performs. The outcomes demonstrate the efficiency of BI-LSTM in handling both forward and backward dependencies from feature sequences to execute sequential data models and to further extract contextual information. Comparisons with various cutting-edge baseline techniques show that the deep learning model is typically more productive and successful in terms of classification quality. Additionally, the model significantly outperforms the findings of the previous models in terms of Accuracy and F1-measure.

In [11], They have put into practice an ensemble model based on the AraBERT and CAME LBERT transformer language models. The balanced dataset, which is made up of reviews of contemporary standard Arabic books, was used to evaluate the suggested ensemble model. Additionally, the proposed model was trained on top of the Twitter dataset, Gold dataset, and ASTD dataset in order to further examine the performance of the model. Compared to the two independent transformer-based models and majority vote. In [12], carried out, an Arabic binary sentiment categorization. They used preprocessing at first to clean up the incoming texts. The LSTM layer has then been fed texts that have been represented as numerical vectors using a word embedding layer. After that, a SoftMax layer was added to predict the text's polarity. The studies had accuracy ranging from 80% to 82%, which were pretty good results.

### 3 Deep learning

Deep learning techniques have created a major breakthrough in artificial intelligence in general and natural language processing in particular. There are many deep learning techniques used to analyze Arabic sentiments, such as the use of CNN techniques and RNN techniques that include LSTM, BI-LSTM, GRU, BI-GRU, and Transformer techniques that have made a major breakthrough in the field of sentiment analysis.

**CNN:** Neural Convolutional Networks CNNs are feedforward neural networks that were initially created for computer vision applications [13, 14] and have demonstrated success in NLP tools[6]. They use a layer that utilizes locally applied convolving filters. Convolution is used instead of generic matrix multiplication, which is a feature of conventional neural networks. It becomes one of the DL algorithms that runs the quickest as a result of the decrease in the number of weights and the consequent decrease in network complexity. CNN furthermore has the benefit of requiring less preprocessing. This opened the way for its use in many other areas, including NLP, voice and handwriting recognition, and picture.

**LSTM:** Long Short-Term Memory (LSTM) networks, a form of recurrent neural network (RNN), are effective at Learning tasks involving sequential input. It resolves these problems by pointing up extensive temporal dependencies. Due to its complexity and module repetition, LSTM is resistant to the optimization issues affecting RNN's basic form [15]. The basic building blocks of the LSTM architecture are a memory cell that maintains its state across time and nonlinear gating devices that control information flow into and out of the cell [16]. Three of the most important gates are input, forget, and output gates. The input block is linked to every gate as well as the output block. The Figure 1 illustrate the component of LSTM framework.

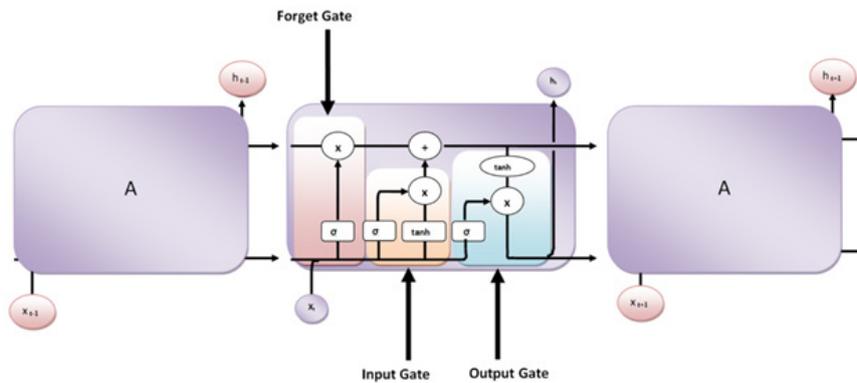


Fig. 1. Component of the LSTM framework

**Bi-LSTM:** A bidirectional LSTM (BiLSTM) layer is used to learn the long-term bidirectional relationships between time steps of time series or sequence data. These dependencies can be useful [17, 18] when you want the network to learn from the full-time series at each time step.

**GRU:** The gated recurrent unit (GRU) framework was proposed in 2014 by [19]. Like LSTM, GRU has gating units that control the flow of information. In GRU, all contents are openly available, in contrast to LSTM networks where the gate limits how much memory may be used by other network nodes. It has been noted that GRU outperforms LSTM in all areas except language modeling [20]. Additionally, the performance gap between the LSTM and GRU networks can be narrowed by initializing the forget gate bias of the LSTM to one. Arabic NLP tasks have already been handled using GRU, notably [21]. The Figure 2 illustrate the component of the GRU framework.

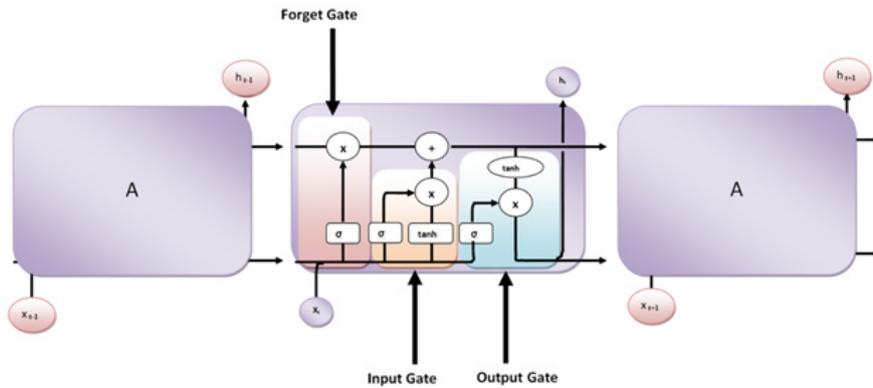


Fig. 2. Component of the GRU framework

**Bi-GRU:** The GRU neural network uses recurrent structures to store and retrieve data over long periods of time, but because it only accesses historical data, its performance may not be as strong in practice as it is in theory [22]. The bidirectional GRU (Bi-GRU) network has a future layer where the data sequence is in the other direction to get around this problem. This network employs two hidden layers that are connected in the output layer in order to harvest information from both the past and the future [23]. These characteristics allow the bidirectional structure to aid the recurrent neural networks in extracting additional information, which increases the efficiency of the learning process [24, 25].

**Transformer:** In [26], the transformer (TRANS) idea was originally put out. The transformer is made comprised of the encoder and decoder's parts. The encoder converts the input sequence into a higher-dimensional space. The output sequence is subsequently generated by the decoder using the mapped input. For translation tasks, it is said to learn far more quickly than recurrent and convolutional systems [26]. With the clear goal of word prediction from context, transformers (feedforward architecture) provide quick training on huge datasets. Even though creating such models is expensive, many of them have previously been made public and are useful for related fields like SA. It has been suggested that certain smaller, supervised datasets may be used to optimize these models.

## 4 Proposed methodology

The proposed system is based on several stages. Firstly, the dataset was chosen. In this paper, the HARD dataset was used. The HARD dataset contains two labels, positive and negative. Secondly, preprocessing the dataset and preprocessing procedures required to get the data ready for the Sentiment analysis task are then discussed. Thirdly, the appropriate embedding was chosen. In this paper, FastText embedding was chosen with all deep learning techniques except AraBERT. Fourth and finally, Various deep learning methods were used to complete the sentiment analysis task. AraBERT technology was used within the transformer and CNN. Two models were worked on, like Alexnet and proposed CNN. And RNN technologies were worked on four types: LSTM, BI-LSTM, GRU, BI-GRU. Finally, an Ensemble Model was made between CNN-RNN and called CNN with BI-GRU model. The Figure 3 illustrates the whole stages for the proposed system, and the deep learning strategies in this work employed in our evaluation are described.

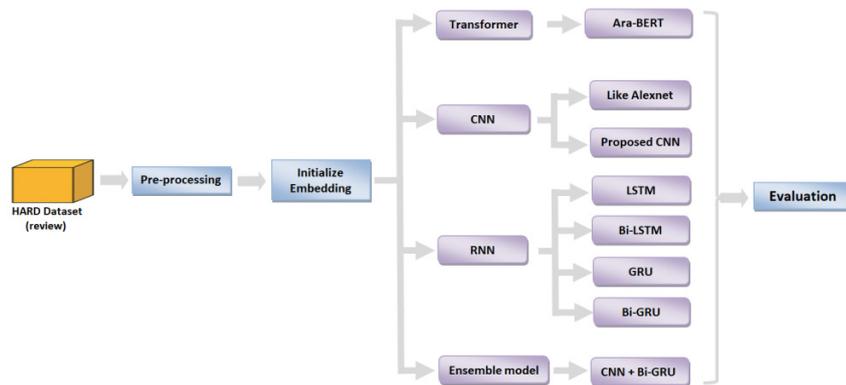


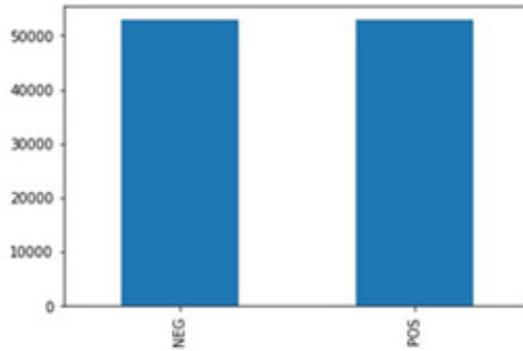
Fig. 3. Proposed methodology

### 4.1 Datasets

HARD. The Hotel Arabic Reviews Dataset (HARD) [27] is the dataset that was used. There are 93700 Arabic-language hotel reviews in this dataset. The hotel reviews were acquired in June and July 2016 from the Booking.com website. The evaluations employ both dialectal Arabic and contemporary standard Arabic. This study will use a balanced dataset (illustrated in Figure 4) with both positive and negative assessments. The ratings are mapped using both positive ratings (4 and 5) and negative ratings (ratings 1 & 2). There are no reviews that are impartial. Table 1 illustrates the number of reviews for the classes. Table 2 shows some reviews of the HARD Dataset.

**Table 1.** Illustrates the stats for the HARD dataset

Review	Number of reviews	Class
1	14382	Negative
2	38467	Negative
4	26450	Positive
5	26399	Positive
All review	105698	



**Fig. 4.** Balanced HARD dataset

**Table 2.** Sample of Hotel Arabic Review Dataset with English Translation

Rating	Review
1	لا انصح". لم يعجبني شيء. لم يعجبني شيء" "I do not recommend." I didn't like it. I didn't like it
2	ضعيف. . عدم الأمان لأني فقدت مبلغ مالي وجوالي ولم يتم التعاون معي weak. . Insecurity because I lost my money and my mobile phone and no cooperation was done with me
4	جيد". ممتاز من جميع النواحي. لا يوجد مواقف للسيارات" "Good". Excellent in all aspects. There is no parking
5	ممتازة". الخدمة والغرفة واسعة والمنظر الرائع والهدوء Excellent". The service and the room are spacious, and the view is wonderful and quiet.

## 4.2 Preprocessing

Reviews contain many words that do not affect the analysis of feelings, whether negative or positive. It is useful to reduce the length of words and thus reduce the size of the word embedding. The data was cleaned and made ready for processing using the following processes.

- Step 1: Read the dataset and check for missing values.
- Step 2: Keep reviewing and rating and drop the rest of the columns.
- Step 3: Mapping each rating value to the specified class by converting the values of 4 and 5 to positive and the values of 1 and 2 to negative.



size of 5, strides of 1, Batch normalization, ReLU activation function, and MaxPooling as its first four sublayers (pool size is 2 and strides is 2). Convolution occurs in the third layer, which receives input from the second layer. There are 384 filters and a 3 kernel size in this layer. The first stride is followed by batch normalization, the ReLU activation function, and finally, the second stride. The fourth layer is also the convolution layer, which receives its inputs from the layer above. This layer has 384 filters, a kernel size of 3, and strides of 1. Batch normalization and ReLU activation function are the next two layers after this layer. The fifth layer is also the convolution layer, which receives its inputs from the preceding layer. This layer has 256 filters, a kernel size of 3, and strides of 1. Batch normalization, ReLU activation function, and MaxPooling layer are all placed after this layer (pool size is 2 and strides is 2). First, of the fully-connected layers, the sixth layer is composed of a Dense layer (4096 units), followed by Batch normalization, the ReLU activation function, and the Dropout layer to prevent overfitting. It receives its input from the flattened layer that follows after the fifth layer. In order to prevent overfitting, the seventh layer additionally comprises a Dense layer (4096 units), Batch normalization, ReLU activation function, and a Dropout layer. In order to prevent overfitting, the eighth layer additionally includes of a Dense layer (1000 units), Batch normalization, ReLU activation function, and a Dropout layer. The softmax activation function is used in the output layer, the last layer. The Alexnet model is shown in Figure 6.

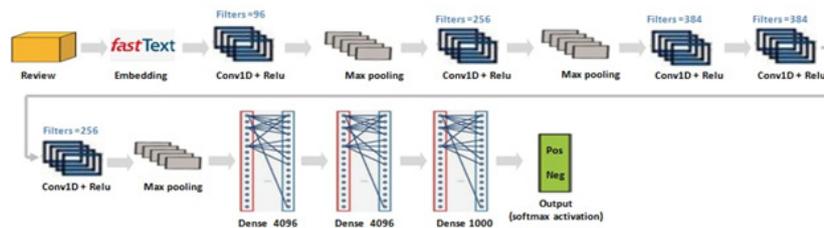


Fig. 6. Proposed model similar to AlexNet

#### 4.4 RNN model

Four RNN Model was built. All models are the same in structure except RNN. LSTM or Bi-LSTM or GRU, or Bi-GRU were used in RNN (only one of them must be used). This model contains the LSTM, Bi-LSTM, GRU, or Bi-GRU layer(Units=256), which takes its input from the Embedding layer, followed by the Dropout layer to avoid overfitting, then followed by the MaxPooling layer (pool size is 2 and strides is 2). To prevent overfitting, the following layer is fully-connected layers(Dense), which are composed of a Dense layer (128 units), a ReLU activation function, and a Dropout layer. Afterward, the Dropout layer was followed by the Dense layer (32 units), which was then followed by the Dense layer (64 units), a ReLU activation function, and finally, the Dropout layer. The last layer is the output layer, which uses the softmax activation. Figure 7 below shows the RNN model.

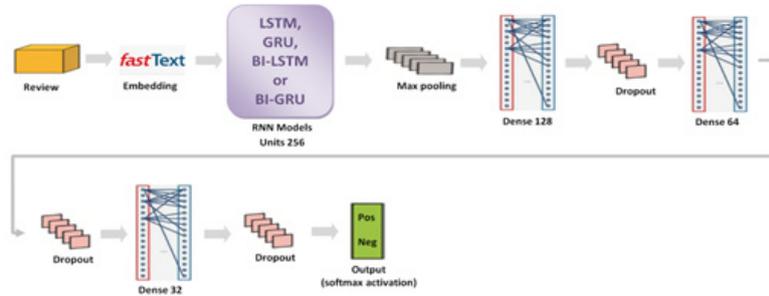


Fig. 7. RNN Model

#### 4.5 Ensemble CNN with BI-GRU model

In this model, BI-GRU is integrated. Where the CNN Model starts with the convolution layer (The number of filters in this layer is 100, the kernel size is 3, and the strides is 1) and then followed by the MaxPooling layer (pool size is 2 and strides is 2) and then followed by the Dropout layer. The next layer is fully-connected layers, and it takes its input from the flattening layer and consists of Dense layer (100 units) and then followed by the Dropout layer. At the same time, the second part consists of the BI-GRU(UNITS=256) layer, followed by the dropout layer. Then the two models merge into one model. Figure 8 shows the Ensemble CNN with the BI-GRU model.

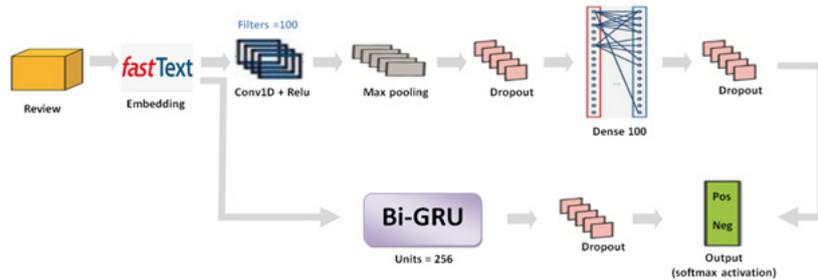


Fig. 8. Ensemble CNN with BI-GRU model

#### 4.6 AraBERT model

In this model was used bert-base-arabertv02-twitter ,Emotional symbols, like emojis, have been introduced to the models' lexicon, along with familiar terms that weren't previously present. AraBERTv0.2-Twitter-base/large are new models for Arabic dialects and tweets that were developed by extending pre-training on about 60 million Arabic tweets utilizing the MLM task (filtered from a collection of 100M). Figure 9 below shows the AraBERT model.



Fig. 9. AraBERT Model

## 5 Experimental results

By calculating the system's accuracy, precision, recall, and f1 score, the system was examined. According to the evaluation of the models, the accuracy, precision, recall, and f1-score for the AraBERT model are all 96.442%, 95.5%, and 97.3%,96.4%, respectively. The confusion matrix is displayed in Figure 10 below (starting from TN, FP, FN, and TP). Table 3 displays the classification report of AraBERT.

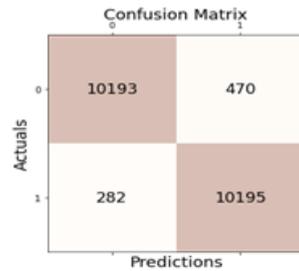


Fig. 10. Confusion matrix (AraBERT)

Table 3. Classification report of AraBERT

	Precision value	Recall value	f1-score value
Negative	0.97	0.96	0.96
Positive	0.96	0.97	0.96
Accuracy	-	-	0.96
macro avg	0.96	0.96	0.96

In CNN, like the AlexNet model, the accuracy is 93.78%, precision is 90.073%, recall is 96.5%, and the f1-score is 93.182%. The confusion matrix is displayed in Figure 11 below (starting from TN, FP, FN, and TP). Table 4 displays the classification report of AlexNet.

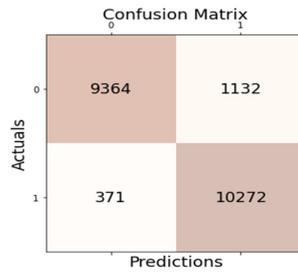


Fig. 11. Confusion matrix (AlexNet)

Table 4. Classification report of CNN Like AlexNet model

	Precision	Recall	f1-score
Negative	0.96	0.89	0.93
Positive	0.90	0.97	0.93
Accuracy	-	-	0.93
macro avg	0.93	0.93	0.93

In the proposed CNN model, the accuracy is 94.43%, and precision is 93.352%, and recall is 95.571%, and the f1-score is 94.448 %. The confusion matrix is displayed in Figure 12 below (starting from TN, FP, FN, and TP). Table 5 displays the classification report of proposed CNN.

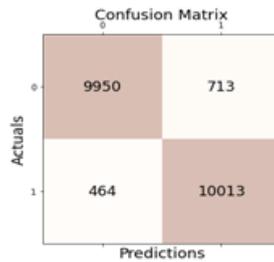
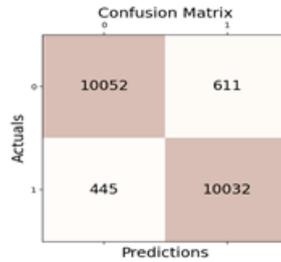


Fig. 12. Confusion matrix (proposed CNN)

**Table 5.** Classification report of proposed CNN model

	<b>precision</b>	<b>Recall</b>	<b>f1-score</b>
Negative	0.96	0.93	0.94
Positive	0.93	0.96	0.94
Accuracy	-	-	0.94
macro avg	0.94	0.94	0.94

In the proposed LSTM model, the accuracy is 95%, and precision is 94.259 %, and recall is 95.752%, and the f1-score is 95%. The confusion matrix is displayed in Figure 13 below (starting from TN, FP, FN, and TP). Table 6 displays the classification report of LSTM.

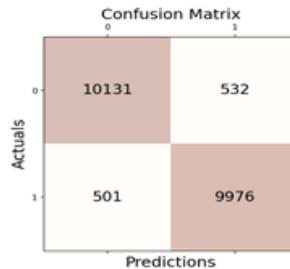


**Fig. 13.** Confusion matrix (LSTM)

**Table 6.** Classification report of proposed LSTM model

	<b>precision</b>	<b>Recall</b>	<b>f1-score</b>
Negative	0.96	0.94	0.95
Positive	0.94	0.96	0.95
Accuracy	-	-	0.95
macro avg	0.95	0.95	0.95

The accuracy, precision, recall, and f1-score of the suggested BI-LSTM model are 95.11%, 94.937%, 95.218%, and 95.077%, respectively. The confusion matrix is displayed in Figure 14 below (starting from TN, FP, FN, and TP). Table 7 displays the classification report of BI-LSTM.

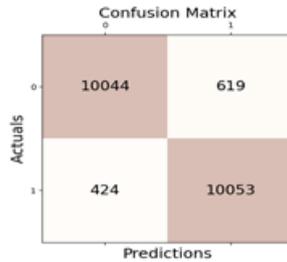


**Fig. 14.** Confusion matrix (BI-LSTM)

**Table 7.** Classification report of proposed BI-LSTM model

	Precision	Recall	f1-score
Negative	0.95	0.95	0.95
Positive	0.95	0.95	0.95
Accuracy	-	-	0.95
macro avg	0.95	0.95	0.95

Accuracy, precision, recall, and f1-score for the proposed GRU model are 95.07%, 94.199%, 95.593%, and 95.068%, respectively. The confusion matrix is displayed in Figure 15 below (starting from TN, FP, FN, and TP). Table 8 displays the classification report of GRU.

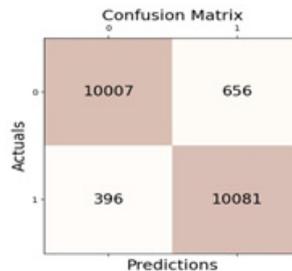


**Fig. 15.** Confusion matrix (GRU)

**Table 8.** Classification report of proposed GRU

	Precision	Recall	f1-score
Negative	0.96	0.94	0.95
Positive	0.94	0.96	0.95
Accuracy	-	-	0.95
macro avg	0.95	0.95	0.95

In the proposed BI-GRU model, the accuracy is 95.02%, precision is 93.89%, recall is 96.22%, and f1-score is 95.041%. The confusion matrix is displayed in Figure 16 below (starting from TN, FP, FN, and TP). Table 9 displays the classification report of BI-GRU.

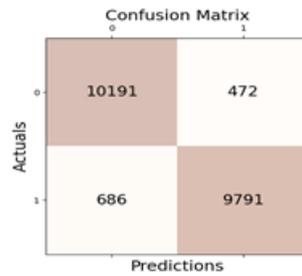


**Fig. 16.** Confusion matrix (BI-GRU)

**Table 9.** Classification report of proposed BI-GRU model

	Precision	Recall	f1-score
Negative	0.96	0.94	0.95
Positive	0.94	0.96	0.95
Accuracy	-	-	0.95
macro avg	0.95	0.95	0.95

In the proposed Ensemble CNN with BI-GRU model, the accuracy is 94.52 %, precision is 95.4%, recall is 93.452%, and f1-score is 94.416%. The confusion matrix is displayed in Figure 17 below (starting from TN, FP, FN, and TP). Table 10 displays the classification report of proposed Ensemble CNN with BI-GRU.



**Fig. 17.** Confusion matrix (Ensemble CNN with BI-GRU)

**Table 10.** Classification report of proposed Ensemble CNN with BI-GRU

	Precision	Recall	f1-score
Negative	0.94	0.96	0.95
Positive	0.95	0.93	0.94
Accuracy	-	-	0.95
macro avg	0.95	0.95	0.95

The results achieved were demonstrated in the different models. The AraBERT has obtained the most accuracy and convergence in precision value, recall value, and f1-score value, RNN, and The AraBERT achieved an accuracy of 96.442%. Then RNN techniques followed it up with accuracy, as the LSTM accuracy reached 95%, the BI-LSTM accuracy reached 95.11%, the GRU accuracy reached 95.07%, and the BI-GRU accuracy reached 95.02%. Ensemble CNN-with-BI-GRU achieved an accuracy of 94.52%. Finally, CNN techniques achieved less accuracy. In the proposed CNN model, the accuracy is 94.43%, while in CNN, like the AlexNet model, the accuracy is 93.78%.

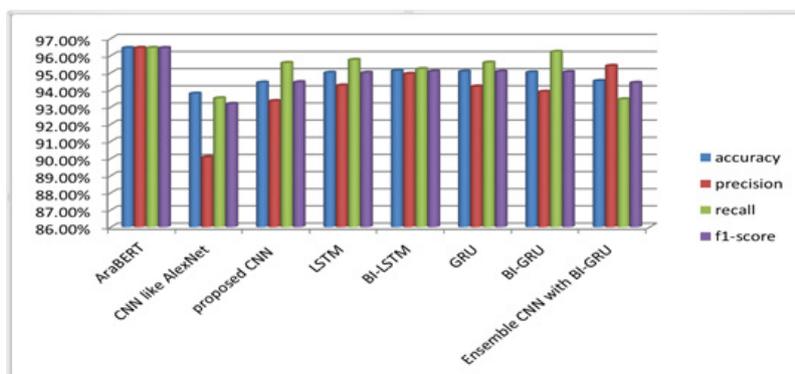


Fig. 18. Comparison of different models

## 6 Conclusion

In this study, a detailed comparison of deep learning methods for Arabic sentiment analysis was carried out. This comparison is the first of its kind because other studies only took some of the strategies this study suggests into account. The most important feature of the fastText representation is that fast and reliably processes big data. Because of the power of fastText embedding, good results have been shown in CNN and RNN techniques. The proposal used Various models such as AraBERT, CNN, and RNN. AraBERT achieved the highest accuracy because it was pre-trained on Arabic language models. The AraBERT attained an accuracy of 96.442%. Then RNN techniques followed it up with accuracy, as the LSTM accuracy reached 95%, the Bi-LSTM accuracy reached 95.11%, the GRU accuracy reached 95.07%, and the Bi-GRU accuracy reached 95.02%. Ensemble CNN-with-Bi-GRU achieved an accuracy of 94.52%. Finally, CNN techniques achieved less accuracy. In the proposed CNN model, the accuracy is 94.43%, while in CNN, like the AlexNet model, the accuracy is 93.78%. The results showed that the like AlexNet model did not achieve high accuracy because the AlexNet was originally designed to process images and was not allocated to texts, but our work used it in the text. In future work, we suggest using an ensemble model between transformer techniques and also suggest adding top layers to AraBERT and freezing some layers that are not very useful, which reduces time.

## 7 References

- [1] B. Brahim, M. Touahria, and A. Tari, "Improving sentiment analysis in Arabic: A combined approach," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 33, no. 10, pp. 1242–1250, 2021. <https://doi.org/10.1016/j.jksuci.2019.07.011>
- [2] J. K. Alwan, A. J. Hussain, D. H. Abd, A. T. Sadiq, M. Khalaf, and P. Liatsis, "Political Arabic articles orientation using rough set theory with sentiment lexicon," *IEEE Access*, vol. 9, pp. 24475–24484, 2021. <https://doi.org/10.1109/ACCESS.2021.3054919>

- [3] A. Khalid Al-Mashhadany, A. T. Sadiq, S. Mazin Ali, and A. Abbas Ahmed, “Healthcare assessment for beauty centers using hybrid sentiment analysis,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 28, no. 2, p. 890, 2022. <https://doi.org/10.11591/ijeecs.v28.i2.pp890-897>
- [4] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, and T. Mikolov, “Learning word vectors for 157 languages,” *arXiv [cs.CL]*, 2018.
- [5] R. I. Farhan, A. T. Maalood, and N. Hassan, “Hybrid feature selection approach to improve the deep neural network on new flow-based dataset for NIDS,” *wjcm*, vol. 1, no. 1, pp. 66–83, 2021. <https://doi.org/10.31185/wjcm.Vol1.Iss1.10>
- [6] A. H. Ombabi, W. Ouarda, and A. M. Alimi, “Deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks,” *Soc. Netw. Anal. Min.*, vol. 10, no. 1, 2020. <https://doi.org/10.1007/s13278-020-00668-1>
- [7] A. Alwehaibi, M. Bikdash, M. Albogmi, and K. Roy, “A study of the performance of embedding methods for Arabic short-text sentiment analysis using deep learning approaches,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 6140–6149, 2022. <https://doi.org/10.1016/j.jksuci.2021.07.011>
- [8] N. Habbat, H. Anoun, and L. Hassouni, “A Novel Hybrid Network for Arabic Sentiment Analysis using fine-tuned AraBERT model,” *Int. J. Electr. Eng. Inform.*, vol. 13, no. 4, pp. 801–812, 2021. <https://doi.org/10.15676/ijeeci.2021.13.4.3>
- [9] A. Mohammed and R. Kora, “Deep learning approaches for Arabic sentiment analysis,” *Soc. Netw. Anal. Min.*, vol. 9, no. 1, 2019. <https://doi.org/10.1007/s13278-019-0596-4>
- [10] H. Elfaik and E. H. Nfaoui, “Deep Bidirectional LSTM Network learning-based Sentiment Analysis for Arabic text,” *J. Intell. Syst.*, vol. 30, no. 1, pp. 395–412, 2020. <https://doi.org/10.1515/jisys-2020-0021>
- [11] I. E. Karfi and S. E. Fkihi, “An ensemble of Arabic transformer-based models for Arabic sentiment analysis,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 8, 2022. <https://doi.org/10.14569/IJACSA.2022.0130865>
- [12] A. Q. Al-Bayati, A. S. Al-Araji, and S. H. Ameen, “Arabic Sentiment Analysis (ASA) using deep Learning approach,” *J. Eng.*, vol. 26, no. 6, pp. 85–93, 2020. <https://doi.org/10.31026/j.eng.2020.06.07>
- [13] Y. Lecun, “A theoretical framework for back-propagation,” in *Proceedings of the 1988 Connectionist Models Summer School*, CMU, Pittsburg, PA, Oxford, England: Morgan Kaufmann, 1988, pp. 21–28.
- [14] J. Q. Kadhim, I. A. Aljazaery, and H. T. H. S. ALRikabi, “Enhancement of online education in engineering college based on mobile wireless communication networks and IOT,” *Int. J. Emerg. Technol. Learn.*, vol. 18, no. 01, pp. 176–200, 2023. <https://doi.org/10.3991/ijet.v18i01.35987>
- [15] J. Schmidhuber, “Deep learning in neural networks: an overview,” *Neural Netw.*, vol. 61, pp. 85–117, 2015. <https://doi.org/10.1016/j.neunet.2014.09.003>
- [16] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, “LSTM: A search space odyssey,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, 2017. <https://doi.org/10.1109/TNNLS.2016.2582924>
- [17] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 13–15 May 2010, vol. 9, pp. 249–256.
- [18] H. T. S. Alrikabi and H. Tuama Hazim, “Secure chaos of 5G wireless communication system based on IOT applications,” *Int. J. Onl. Eng.*, vol. 18, no. 12, pp. 89–105, 2022. <https://doi.org/10.3991/ijoe.v18i12.33817>
- [19] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio, “On the properties of neural machine translation: Encoder–decoder approaches,” in *Proceedings of SSST-8*, Eighth

- Workshop on Syntax, Semantics and Structure in Statistical Translation, 2014. <https://doi.org/10.3115/v1/W14-4012>
- [20] R. Jozefowicz, W. Zaremba, and I. Sutskever, “An empirical exploration of recurrent network architectures,” in Proceedings of the 32nd International Conference on Machine Learning, 07–09 Jul 2015, vol. 37, pp. 2342–2350.
- [21] S. Al-Azani and E.-S. El-Alfy, “Emojis-based sentiment classification of Arabic microblogs using deep recurrent neural networks,” in 2018 International Conference on Computing Sciences and Engineering (ICCSE), 2018. <https://doi.org/10.1109/ICCSE1.2018.8374211>
- [22] Y. Deng, H. Jia, P. Li, X. Tong, X. Qiu, and F. Li, “A deep learning methodology based on bidirectional gated recurrent unit for wind power prediction,” in 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2019. <https://doi.org/10.1109/ICIEA.2019.8834205>
- [23] X. Luo, W. Zhou, W. Wang, Y. Zhu, and J. Deng, “Attention-based relation extraction with bidirectional gated recurrent unit and highway network in the analysis of geological data,” IEEE Access, vol. 6, pp. 5705–5715, 2018. <https://doi.org/10.1109/ACCESS.2017.2785229>
- [24] D. Zhang, L. Tian, M. Hong, F. Han, Y. Ren, and Y. Chen, “Combining convolution neural network and bidirectional gated recurrent unit for sentence semantic classification,” IEEE Access, vol. 6, pp. 73750–73759, 2018. <https://doi.org/10.1109/ACCESS.2018.2882878>
- [25] A. Saleh Hussein, R. Salah Khairy, S. M. Mohamed Najeeb, and H. T. S. Alrikabi, “Credit card fraud detection using fuzzy rough nearest neighbor and sequential minimal optimization with logistic regression,” Int. J. Interact. Mob. Technol., vol. 15, no. 05, p. 24, 2021. <https://doi.org/10.3991/ijim.v15i05.17173>
- [26] A. Vaswani et al., “Attention is all you need,” arXiv [cs.CL], 2017.
- [27] A. Elnagar, Y. S. Khalifa, and A. Einea, “Hotel Arabic-reviews dataset construction for sentiment analysis applications,” in Intelligent Natural Language Processing: Trends and Applications, Cham: Springer International Publishing, 2018, pp. 35–52. [https://doi.org/10.1007/978-3-319-67056-0\\_3](https://doi.org/10.1007/978-3-319-67056-0_3)

## 8 Authors

**Rafeef Abd Al-Ameer** obtained a Bachelor's degree in Computer Science from the University of Baghdad in 2014. She is a M.Sc student at the University of Technology (UOT) – Iraq.

**Wael J. Abed**, Prof. Dr. in Computer Techniques Engineering Department, Al-Mustaqbal University College.

**Ahmed T. Sadiq** is a Professor in the Computer Science Department university of Technology Iraq. He received a B.Sc., M.Sc. & Ph.D. degree in Computer Science from the University of Technology.

Article submitted 2023-02-10. Resubmitted 2023-03-28. Final acceptance 2023-03-29. Final version published as submitted by the authors.