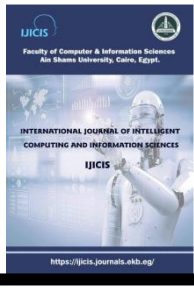
IJICIS, Vol.24, No.4, 99-107
DOI: 10.21608/ijicis.2025.348244.1370

International Journal of Intelligent Computing and Information Sciences

https://ijicis.journals.ekb.eg/



ARABIC EMOTION-BASED SENTIMENT ANALYSIS USING ENSEMBLE DEEP LEARNING MODEL

Ali Abdallah
Information Systems Department, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt
20201700504@ cis.asu.edu.eg

Mohamed Ali
Information Systems Department, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt
20201701233@ cis.asu.edu.eg

Abdulrhman Abdelhalim
Information Systems Department, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt
20201700442@ cis.asu.edu.eg

Ali Maher
Information Systems Department, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt
20201700506@ cis.asu.edu.eg

Karima Sobhi
Information Systems Department, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt
20201701115@ cis.asu.edu.eg

Hany Mohamed
Information Systems Department, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt
20201700951@ cis.asu.edu.eg

Shaimaa Haridy*
Information Systems Department, Faculty of Computer and Information Sciences, Ain Shams University Cairo, Egypt
shaimaaharidy@cis.asu.edu.eg

Received 2024-12-28; Revised 2025-01-22; Accepted 2025-01-22

Abstract: *In today's digital age, social media platforms have become a prevalent medium for individuals to share their opinions, emotions, and experiences. Despite the surge in user-generated content, effective tools and resources for sentiment analysis in the Arabic language remain insufficient. This paper addresses this gap by presenting a novel approach to Arabic sentiment analysis through the development of a web application for text emotion classification. The proposed methodology employs an ensemble of deep learning models, including Bidirectional Long Short-Term Memory (Bi-LSTM), Bidirectional Gated Recurrent Unit (Bi-GRU), and the MARBERTv2 transformer model, combined using a Random Forest stacking technique. The system's performance is evaluated on the Emotone_ar dataset, providing a robust benchmark for emotion detection tasks. Experimental results demonstrate that the ensemble model outperforms individual models, achieving an accuracy of 90%, a recall of 90%, and an F1 score of 90%. The integration of MARBERTv2, a pre-trained language model specifically designed for Arabic, shows superior performance compared to other models tailored for the Arabic language. This work concludes that the proposed ensemble model not only advances the field of*

***Corresponding Author**: Shaimaa Haridy

Information Systems Department, Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt

Email address: shaimaaharidy@cis.asu.edu.eg

Arabic sentiment analysis but also offers an effective tool for real-time emotion detection in social media texts, addressing a critical need in natural language processing for Arabic.

Keywords: *Arabic sentiment analysis, emotion classification, deep learning, social media analysis*

1. Introduction

In today's digital age, the proliferation of textual data across social media, forums, and other online platforms has created a rich resource for understanding human emotions. Accurately detecting and analyzing these emotions can provide valuable insights into public opinion, customer satisfaction, and social trends. This is particularly important for Arabic text, given the language's complex morphology and diverse dialects, which pose significant challenges for natural language processing (NLP).

To address this need, we have developed a web application for Arabic sentiment analysis that focuses on emotion detection. The proposed application leverages a sophisticated model combining Bidirectional Gated Recurrent Units (BI GRU), Bidirectional Long Short-Term Memory (BI LSTM), MARBERT, and Random Forest stacking to achieve an impressive 90% accuracy. This high level of accuracy ensures reliable emotion detection for various applications, from social media monitoring to market research.

The proposed application offers two primary functionalities. Firstly, users can upload an Excel sheet containing textual data. The application processes this data, performs sentiment analysis, and generates a comprehensive dashboard that details the emotional landscape of the data. Users can also download an updated Excel sheet with an added column for emotion classification, facilitating further analysis and reporting.

Secondly, the proposed application allows users to search for specific words. It retrieves the 50 most recent tweets related to the search term from Twitter, performs sentiment analysis, and presents the findings in a user-friendly dashboard. This feature is particularly useful for real-time monitoring of public sentiment on trending topics.

The proposed work utilizes the Emotone_ar dataset, which is specifically designed for Arabic emotion detection. The importance of emotion detection in textual data cannot be overstated. It enables organizations to harness the vast amounts of textual data generated daily, turning it into actionable insights. By focusing on emotion detection, this application helps bridge the gap between raw data and meaningful interpretation, empowering users to make informed decisions based on emotional trends.

Overall, this web application provides a robust and accurate tool for Arabic sentiment analysis, with a particular emphasis on emotion detection. Its dual functionalities cater to both batch processing of large datasets and real-time analysis of social media trends, making it a versatile solution for various analytical needs. This work not only advances the field of Arabic NLP but also offers practical tools that can significantly impact areas such as customer feedback analysis, social media management, and market research.

2. Related Work

Extensive research has been conducted to derive emotional insights and comprehend the context of English text. However, relatively few studies focus on the Arabic language due to several challenges, such as its linguistic complexity, the limited availability of Arabic resources, and the variety of Arabic dialects. Recently, there has been a growing interest in analyzing the Arabic language, driven by the increasing need for digital transformation in Arab communities. Arabic, spoken in many different countries, presents significant analytical challenges due to its dialectical diversity. Table 1 provides a comparative summary of recent related research.

TABLE 1. Comparison between related works

Reference	Year	Problem	Dataset	Methodology
User satisfaction with Arabic COVID-19 apps: Sentiment analysis of users' reviews using machine learning techniques [5]	2024	Arabic sentiment analysis of users' reviews	COVID-19 Apps benchmark dataset	SVM KNN NB LR RF ANN
Improved Emotion Detection Framework for Arabic Text using Transformer Models [6]	2023	Arabic Single-label Emotion Classification	Emotone_ar	Arabic BERT-based Model
An Ensemble Deep Learning Approach for Emotion Detection in Arabic Tweets [1]	2022	Arabic Multilabel Emotions Classification	SemEval2018 task 1- Ec-Ar	MARBERT, Bi-LSTM, Bi-GRU
Combining Context-aware Embeddings and an Attentional Deep Learning Model for Arabic Affect Analysis on Twitter [3]	2021	Arabic Multilabel Emotions Classification	SemEval2018 task 1- Ec-Ar	AraBERT word embeddings, attention-based LSTM and Bi-LSTM
Hybrid Feature Model for Emotion Recognition in Arabic Text [2]	2020	Arabic Multilabel Emotions Classification	SemEval2018 task 1- Ec-Ar	HEF + DF Hybrid of human-engineered feature-based model +deep feature-based (DF) model
Arabic dialect sentiment analysis with ZERO effort [4]	2020	Arabic dialect's sentiment analysis without constructing any resources	TSAC ElecMorocco2016	MLP CNN LSTM bi-LSTM

After reviewing and studying the papers and conducting a search to select the best model to detect and classify emotions from Arabic texts, we conclude that the best technique to use is ensemble deep learning approach by combining strong deep learning models such as Bi-LSTM, Bi-GRU and MARBERT.

3. Proposed System

The proposed system for Arabic emotion detection is designed as a web application that classifies text emotions using an ensemble of deep learning models. The architecture is divided into three layers: the Presentation Layer, the Business Logic Layer, and the Data Layer.

3.1. System Architecture

This entire process of the proposed system is depicted in Fig. 1. In the next subsections the three layers are explained in detail.

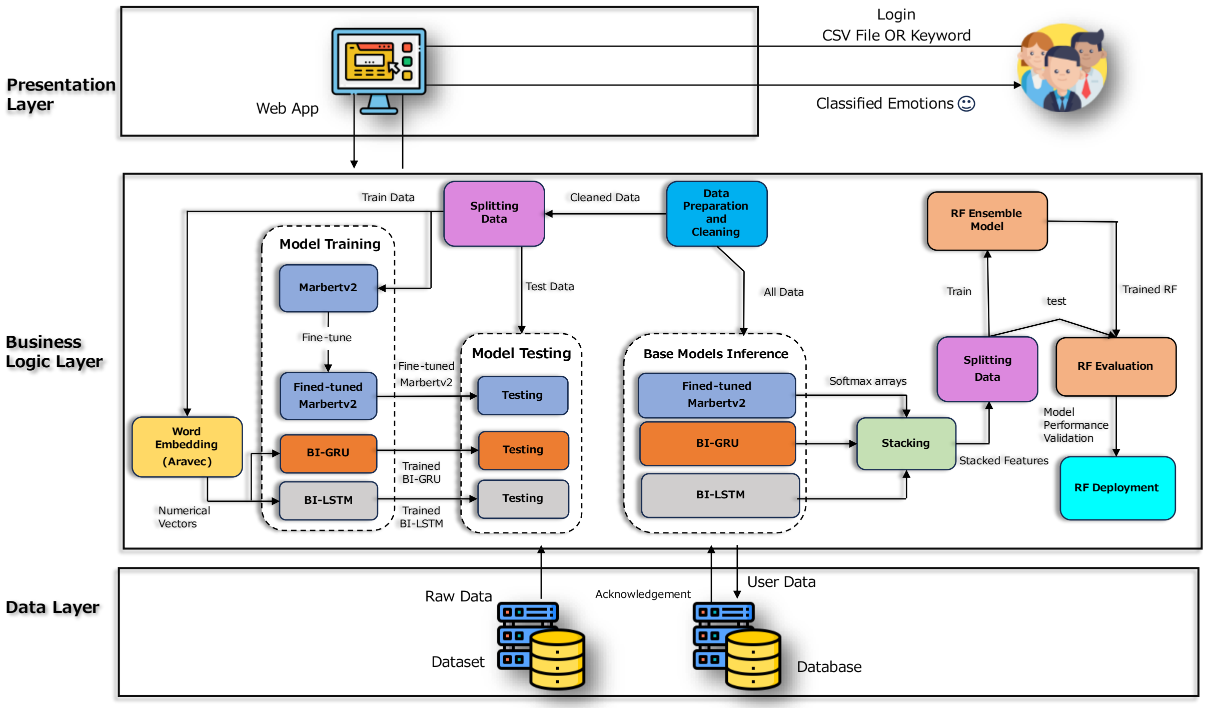


Figure. 1: System Architecture

3.2. Presentation Layer

The web application interface allows users to log in and classify emotions in text by uploading CSV files or inputting keywords. Users can upload CSV files containing tweets or reviews for comprehensive sentiment analysis. Alternatively, they can search Twitter by entering specific keywords to retrieve the latest 50 tweets containing those keywords.

The results are displayed on an interactive dashboard with detailed charts, graphs, and tables, providing insightful analysis of sentiments and trends in the processed text.

3.3. Business Logic Layer

This layer serves as the foundation of our platform. It is divided into the next modules.

3.3.1. Data Preparation and Cleaning

This module cleans the input data, removing null and duplicated entries to improve models' accuracy. At the preprocessing step we used the following common methods:

- Remove Emails: Deletes any email addresses present in the text.
- Remove URLs: Deletes URLs from the text.
- Remove Mentions: Deletes Twitter mentions (usernames).
- Hashtags to Words: Converts hashtags into readable words.
- Remove Punctuations: Deletes punctuation marks.
- Normalize Arabic: Normalizes Arabic text for consistent representation.
- Remove Diacritics: Deletes diacritical marks.
- Remove Repeating Characters: Deletes characters that repeat excessively.
- Remove Newlines: Deletes newline characters.
- Remove Stop Words: Deletes common Arabic stop words.
- Replace Emojis: Emojis reflect a tweeter's feelings, and their existence is very important for understanding the context, so we have replaced them with plain text because it may affect the classification of tweets.
- Remove English Characters: Deletes any English letters.
- Remove Digits: Deletes digits.

Table 2 shows examples of tweets before and after preprocessing:

Table 2: Comparison between original and proceed tweets after preprocessing step

Original Tweet	Processed Tweet
ابيه المباراة دي!! 😠	ايه المباراة دي غضب
انا بجد تعبت وزهقت كل شويه انسي اللي ذاكرته.. 💔 😰	بجد تعبت وزهقت شويه انسي اللي ذاكرته.. حزن حزن
@karima ❤️ نحب ماتشات الكرة بزاف	نحب ماتشات الكرة بزاف قلب

3.3.2. Word Embedding (AraVec)

Converts raw text into numerical vectors suitable for Bi-LSTM and Bi-GRU model training.

3.3.3. Model Training

Utilizes MARBERTv2 for initial training, which is then fine-tuned. Bi-GRU and Bi-LSTM models are also trained.

3.3.4. Model Testing

Evaluates the performance of each trained model.

3.3.5. Base Models Inference

Generates SoftMax arrays from fine-tuned MARBERTv2, Bi-GRU, and Bi-LSTM models.

3.3.6. Stacking

Combines outputs from the base models to create a feature set for the Random Forest (RF) ensemble model.

3.3.7. Random Forest Ensemble Model

Trains and evaluates the ensemble model using the stacked features. The trained model is then deployed for real-time emotion classification.

3.4. Data Layer

The dataset contains raw data, which is then cleaned and split into training, validation, and testing sets for model development. Specifically, 80% of the data is allocated for training, 10% for validation and 10% for testing.

3.5. Workflow

- Data Input: Users log in to the web app and provide data either by uploading CSV files or entering keywords.
- Data Preparation: The data undergoes preprocessing and cleaning to ensure quality.
- Model Training and Testing: Deep learning models (MARBERTv2, Bi-GRU, Bi-LSTM) are trained and tested. The fine-tuned MARBERTv2 model is integrated with other models.
- Model Inference and Stacking: Outputs from individual models are stacked to create a robust feature set.
- Ensemble Model: The Random Forest model uses stacked features to classify emotions, which are then returned to the user.

  This architecture ensures efficient emotion classification, leveraging advanced deep learning techniques to enhance the accuracy and reliability of the system.

4. Expermintal Results

In this section, dataset and results will be explained.

4.1. Dataset

The 'Emotone_ar' dataset is a comprehensive collection of 10,065 Arabic tweets, annotated with eight emotions: sadness, anger, joy, surprise, love, sympathy, fear, and none. This dataset was created by researchers at Nile University in Egypt and first published in 2017. Tweets were collected using emotion-related keywords and manually annotated by three native Arabic speakers. Emotone_ar is a multi-dialect dataset, including various Arabic dialects to ensure broad applicability. The balanced and multi-dialect nature of the Emotone_ar dataset supports robust model training, making the proposed system effective for various Arabic dialects and emotional expressions.

As displayed in Table 3, the dataset is balanced, with nearly equal representation of each emotion, making it suitable for training and evaluation

Table 3: Number of Tweets for Emotone_ar After Cleaning

Emotion	Count
None	1539
Anger	1440
Fear	1204
Sadness	1254
Joy	1280
Love	1213
Sympathy	1046
Surprise	1044
Total	10,024

4.2. Results

The proposed system has been evaluated on the Emotone_ar dataset. The experimental results demonstrate that the ensemble model outperforms individual models, achieving an accuracy of 90%, a recall of 90%, and an F1 score of 90%. Table 4 summarizes a comparison with related works.

Table 4: Results Comparison with related works

Paper	Performance Metrics
[5]	Accuracy: 0.89 F1-score: 0.89
[6]	Accuracy: 0.74 F1-score: 0.74
[1]	Accuracy: 0.540 Macro F1 Score: 0.701 Recall: 0.550 Micro F1 Score: 0.527
[3]	Accuracy: 0.538
[2]	Micro F1: 0.631 Macro F1: 0.502 Jaccard Acc: 0.512
[4]	F1-score: 0.83

5. Conclusion and Future Work

In this study, we developed a web application for Arabic sentiment analysis, achieving significant advancements in natural language processing (NLP) tailored to Arabic text. By employing sophisticated models such as Bidirectional Gated Recurrent Units (Bi-GRU), Bidirectional Long Short-Term Memory (Bi-LSTM), MARBERT, and Random Forest in a stacking ensemble approach, we achieved a high accuracy rate of 90%. This demonstrates the efficacy of our methodology in accurately detecting and classifying a wide range of emotions, including love, anger, fear, and more.

Our application is designed with user convenience and versatility in mind, offering two primary functionalities. Users can upload Excel sheets containing textual data for comprehensive analysis, complete with a detailed dashboard and an option to download the enriched dataset with emotion classifications. Additionally, users can search for specific keywords to retrieve and analyze the latest 50 tweets from Twitter, providing real-time insights into public sentiment. The resulting dashboards present a clear and intuitive visualization of the sentiment distribution, enhancing data interpretation and utilization.

Future research should focus on expanding and diversifying the dataset to include various forms of Arabic text from multiple sources and dialects to create a more comprehensive training corpus. Incorporating contextual information into the sentiment analysis model through context-aware NLP models and additional tweet metadata is another important direction. Leveraging advanced NLP techniques, such as transformers and attention mechanisms, will better enable understanding of nuanced language features like sarcasm and irony. Developing models tailored to specific Arabic dialects can improve accuracy and relevance across different Arabic-speaking regions. Integrating the sentiment analysis model into real-time monitoring systems will help handle high volumes of data with minimal latency. Implementing mechanisms for user feedback to continuously improve the model through active learning and periodic retraining on newly annotated data will also be beneficial.

Additionally, integrating the AraT5 model could further enhance the model's performance in understanding and generating Arabic text. Expanding the application's capabilities to include a Facebook keyword search feature would provide a broader scope of social media sentiment analysis. Refining the user interface will improve user experience and accessibility, making the application more intuitive and user-friendly. By addressing these areas, we aim to further enhance the accuracy, reliability, and applicability of our Arabic sentiment analysis web application, making it a more powerful tool for understanding and analyzing emotions in Arabic text

References

1. Alaa Mansy, Sherine Rady, and Tarek Gharib. "An ensemble deep learning approach for emotion detection in arabic tweets." International Journal of Advanced Computer Science and Applications 13.4 (2022).
2. N. Alswaidan and M. E. B. Menai, "Hybrid Feature Model for Emotion Recognition in Arabic Text," IEEE Access, vol. 8, pp. 37843-37854, 2020, doi: 10.1109/ACCESS.2020.2975906.
3. H. Elfaik and E. H. Nfaoui, "Combining Context-Aware Embeddings and an Attentional Deep Learning Model for Arabic Affect Analysis on Twitter," IEEE Access, vol. 9, pp. 111214-111230, 2021, doi: 10.1109/ACCESS.2021.3102087.
4. Guellil, I., Mendoza, M., & Azouaou, F. (2020). Arabic dialect sentiment analysis with ZERO effort.\\Case study: Algerian dialect. Inteligencia Artificial, 23(65), 124-135..
5. Ramzy, M., & Ibrahim, B. (2024). User satisfaction with Arabic COVID-19 apps: Sentiment analysis of users' reviews using machine learning techniques. Information Processing & Management, 61(3), 103644..
6. H. Noaman, "Improved Emotion Detection Framework for Arabic Text using Transformer Models," Advanced Engineering Technology and Application, vol. 12, no. 2, pp. 1-11, 2023.
7. A. Alqarni and A. Rahman, "Arabic Tweets-based Sentiment Analysis to investigate the impact of COVID-19 in KSA: A deep learning approach," Big Data and Cognitive Computing, vol. 7, no. 1, p. 16, 2023.

8. M. Abdul-Mageed, A. Elmadany, and E. M. B. Nagoudi, "MARBERT: Deep Bidirectional Transformers for Arabic," in Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 7088-7105, 2021. Online. DOI: 10.18653/v1/2021.acl-long.551.
9. A. B. Soliman, K. Eisa, and S. R. El-Beltagy, “AraVec: A set of Arabic Word Embedding Models for use in Arabic NLP,” in Proceedings of the 3rd International Conference on Arabic Computational Linguistics (ACLing 2017), Dubai, UAE, 2017.
10. A. Al-Khatib and S. El-Beltagy, "Emotional Tone Detection in Arabic Tweets," in Proceedings of the 18th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2017), Budapest, Hungary, 2017.