Textual Arabic Emotion Detection Using a Hybrid Approach with Generative Kernels

Nuha Zamzami[1] and Fatma Najar[2]

Abstract— **Emotion detection has emerged as an interesting topic for researchers due to its importance in marketing, advertising, social behavior analysis, and human-computer interaction. Machine learning and deep learning algorithms have proven to be efficient for detecting emotions in textual data from social media. However, the majority of studies in the literature focus on English conversations, with very few efforts addressing the same issue in other languages. With the increased volume of Arabic social media content, automated emotion detection in Arabic has become essential. This work investigates emotion detection in Arabic text by proposing a hybrid approach combining deep learning models and Support Vector Machines with Fisher kernels trained via unsupervised learning. The experiments were evaluated on the Arabic Emotion Twitter dataset, one of the largest annotated Arabic datasets with over $10K$ tweets and eight diverse emotions. Experiments demonstrated that the proposed approach achieved superior performance compared to previous work, achieving an overall precision of up to 86%. Furthermore, it outperforms the standalone approaches and the recent, closely related work on the same dataset, achieving accuracy improvements of around $24-25\%$ and 6%, respectively.**

Arabic Emotion detection, Hybrid approach, Supervised learning models, Deep learning, Unsupervised learning models, Fisher kernels.

I. Introduction

Social media platforms are valuable sources of information, enabling users to engage in discussions and share their thoughts and opinions. Users' opinions are typically subjective expressions of their attitudes, assessments, or emotions regarding organizations, events, and products. While sentiment analysis focuses on the polarity of opinions (positive or negative), emotion detection goes deeper to explore and describe individuals' emotions [1]. Textual emotion detection has applications in numerous aspects of daily life, including the development of efficient e-learning systems tailored to students' emotions, the enhancement of human-computer interactions, adjusting business strategies based on customer emotions, identifying potential criminals or terrorists by analyzing emotional responses, and optimizing the performance of chatbots and other automated feedback systems [2].

The analysis of emotions in social media presents numerous challenges due to prevalent grammatical errors, misspellings, slang, social shortcuts, and multimedia content. Multiple studies in the literature analyze emotion in English-language tweets (e.g., [3]), while few focus on Arabic-language tweets (see, for example, [4],[5]). The Arabic language still lacks studies, datasets, and resources in the field of emotion detection. Researchers explore different approaches for detecting emotion in text, including keyword-based, lexical-based, machine learning, deep learning, transformer models, and hybrid approaches [6]. The hybrid approach is useful when standalone approaches fail to deliver satisfactory results. A hybrid system uses a combination of two or more different approaches to achieve enhanced performance. A few recent works proposed hybrid models to detect emotion in Arabic text. For instance, in [7], the authors proposed a hybrid model that combined a Convolutional Neural Network (CNN) and a Bidirectional Gated Recurrent Unit (Bi-GRU) as deep learning techniques, and Support Vector Machines (SVM) as a supervised learning approach. Owing to their computational efficiency and superior discrimination skills, SVMs have emerged as a standard supervised learning method with benchmark results.

In this work, we extend the research in [7] by proposing a hybrid approach that combines deep learning and SVM with generative kernels, aiming to enhance emotion recognition performance and be adapted to Arabic text. In SVM, if the data is not linearly separable, it can be transformed into a high-dimensional feature space utilizing a kernel function to facilitate the computation of the inner product of the changed data in that space [8]. However, choosing a suitable SVM kernel presents a significant challenge in machine learning and data mining applications. Classic kernels fail to account for the data's inherent characteristics. An improved method involves generating kernels directly from the data, also known as generative kernels. The Fisher kernel, introduced in [9], exemplifies earlier attempts to develop more adaptable kernels and has been widely used in the literature. Recent initiatives have focused on integrating the benefits of both methodologies by creating hybrid generative/discriminative algorithms (see, for example, [10], [11]). These algorithms have demonstrated efficacy as robust tools that often yield an enhanced performance compared to fully generative or discriminative methods.

The rest of this work is organized as follows. Section II presents a brief review of previous efforts to detect emotions in Arabic text. Then, the proposed system is explained in detail in section 2. The section after IV is devoted to the experimental results and analysis. Finally, the last section V concludes this study along with recommendations for future research.

[1]Nuha Zamzami is with the Department of Computer Science and Artificial Intelligence, College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia. nezamzami@uj.edu.sa

[2]Fatma Najar is with the Department of Mathematics and Computer Science, John Jay College of Criminal Justice, The Graduate Center, The City University of New York, New York, NY, USA. fnajar@jjay.cuny.edu

II. Related Work

Al-Khatib and El-Beltagy [12] proposed a machine learning method for predicting the emotions of Arab users on Twitter. They published a new dataset based on Twitter (rebranded as X) that aims to cover the most frequently used emotion categories. The data was gathered from different sources and then annotated and cleaned by removing hashtags, advertisements, and duplicates. The authors represent the dataset using a bag-of-words (BOW) approach, which they use as input to different classifiers. The best results were obtained using a complementary Naïve Bayes algorithm, which had an overall accuracy of 68.12%. AlFutamani and Al-Baity [13] used supervised machine learning classifiers to detect emotions in Arabic tweets (primarily Saudi-based). The SVM and logistic regression classifiers achieved the best performance, with an overall accuracy of 73.39%. Al-Hagery et al. [14] investigated the effects of feature extraction techniques and N-grams on the performance of three supervised machine learning algorithms: SVM, LR, and Naive Bayes (NB). The dataset used in this study was a real-world Twitter dataset containing $3,171$ tweets that were collected using the Netlytic tool. The SVM classifier achieved the best performance in classifying emotions in Arabic tweets, with an accuracy of 82.43% using a BoW weighting scheme.

Abdullah and Shaikh [15] proposed a novel approach to detect emotions in English and Arabic tweets based on the SemEval-2018 dataset. The authors applied a fully connected neural network architecture in which the main input to the system is a combination of word2vec and doc2vec embeddings, along with a set of psycholinguistic features (e.g., from the AffectiveTweets Weka package). The proposed deep learning approach achieved an accuracy of 59.70% for detecting the intensity of emotion in Arabic. Mansy et al. [4] examined the performance of emotion detection in Arabic tweets using an ensemble deep learning method. The proposed model is based on Bi-LSTM, Bi-GRU, and the MARBERT transformer. The SemEval-2018 dataset was used to evaluate the proposed ensemble model. The ensemble model achieved an accuracy of 54%, outperforming the individual models. Bharti et al. [7] introduced a hybrid (machine learning and deep learning) model to detect emotions in Arabic text using TF-IDF and word2vec embeddings. The architecture based on a combined CNN, GRU, and SVM outperforms both machine learning and deep learning methods, achieving 80.11% accuracy on the SemEval-2018 dataset.

However, in the above-mentioned approach, the architecture used classical SVM kernels, which are too simple to capture the complexity of Arabic text features. Choosing the right kernels is a crucial step in building an SVM classifier because it significantly affects the model's performance and robustness.

III. Proposed Methodology

This section outlines the proposed Arabic emotion recognition framework, including data description and preprocessing. After preprocessing, the data will be transformed into a vector representation using word embeddings and provided to deep learning models. In addition to evaluating individual models, we combine the two superior deep learning models based on accuracy and F1 score. The integration of these high-level features of textual data yields a latent vector, which serves as input to the SVM using generative kernels for emotion detection.

A. Dataset Description and Preprocessing

In this work, we use the Twitter-balanced Arabic emotion dataset presented in [12] [1]. The dataset contains over 10k tweets aggregated from multiple sources. It covers the most frequently used emotion categories in Arabic tweets: joy/happiness ($1,281$ tweets), anger ($1,444$ tweets), sympathy ($1,062$ tweets), sadness ($1,256$ tweets), fear ($1,207$ tweets), surprise ($1,045$ tweets), love ($1,220$ tweets), and none ($1,550$ tweets), i.e., tweets with neutral emotions (none label). Samples of data in each class are shown in Figure 1.

The first part of our proposed approach is extracting the relevant features from the Arabic Emotion Twitter dataset. Feature extraction involves four steps as follows:

- Normalization: The first phase is to normalize Arabic characters and, Hindi numerals were also replaced by Arabic numerals.
- Diacritics Removal: In this step, Arabic diacritics were removed.
- Punctuation Removal: Removing all punctuation marks.
- Hyperlinks, mentions starting with @, emojis, or retweet signs (RT) were removed.

After extracting features, we vectorize them using various word embedding techniques, including Word2Vec, Embeddings from Language Models (ELMo), FastText, and Global Vectors for Word Representation (GloVe). This process generates an embedding layer of size $(15098, 300)$, which will serve as the input for our hybrid approach. After performing these steps, the data is split into 70% training, 20% validation, and 10% test sets using the different deep learning algorithms.

B. The Hybrid Arabic Emotion Detection Model

The proposed hybrid model combines supervised learning models, including deep learning and SVM models, with generative kernels derived from supervised learning models (finite GMM and infinite GMM) to predict emotions. The overall Hybrid Arabic Emotion detection diagram is illustrated in Figure 2.

1) Deep Learning Models: The first stage of our experiments includes the investigation of state-of-the-art deep learning models, namely CNN, GRU, Bi-GRU, LSTM, and BiLSTM, using different embedding techniques. Below, we briefly explain each of the deployed models:

- **Convolutional Neural Network (CNN):** CNN is a class of neural networks capable of extracting position-invariant and local characteristics. CNN accepts d-dimensional dense vectors, or word vectors, as input,

[1] The dataset is publicly available and can be downloaded from https://huggingface.co/datasets/emotone-ar-cicling2017/emotone$_{a}r$

No.	Text	Label
1	خيبة أمل مصريه في الاوليمبياد يتحملها اتحاد فاشل بكل المقاييس Egypt's disappointment in the Olympics is the fault of a failed federation by all standards.	غضب Anger
2	ليش احس اني الوحيدة اللي ميته خوف من نتايج انتخابات أمريكا Why do I feel like I'm the only one who's scared to death of the US election results?	خوف Fear
3	حلب تنتصر خبر سعيد جداً الله يكثر الأخبار الحلوة ملحمة حلب الكبرى Aleppo is victorious, very happy news. May God increase the good news. The great epic of Aleppo.	مرح Joy
4	اشتقت لشخص كان لديه القدرة على تغير مزاجي السيء .. كلام بوح عابر اشتياق القلب I miss someone who had the ability to change my bad mood.. fleeting words of confession, longing of the heart.	حب Love
5	مش في الاولمبياد لكن في مكان مطلوب فيه تحديداً الزي الوطني Not in the Olympics, but in a place where the national dress is specifically required.	غير محدد None
6	الصراحة إحساس مخزي ونحن نكتفي بالتمثيل المشرف فقط في الاولمبياد عاوزين صحوة رياضيه Frankly, it is a shameful feeling that we are satisfied with only honorable representation in the Olympics. We want a sports awakening.	حزن Sadness
7	انتو متخيلين ان في خيول بتسافر ريو دي جانيرو علشان تسابق في الاولمبياد واحنا قاعدين هنا؟ Can you imagine that horses are traveling to Rio de Janeiro to compete in the Olympics while we are sitting here?	مفاجأة Surprise
8	تقلب الأجواء واشتداد البرودة يتعب الفقراء الله يكون بعونهم The changeable weather and the intense coldness are tiring for the poor. May God help them.	تعاطف Sympathy

Fig. 1. A Sample of the Emotional Tone Dataset.

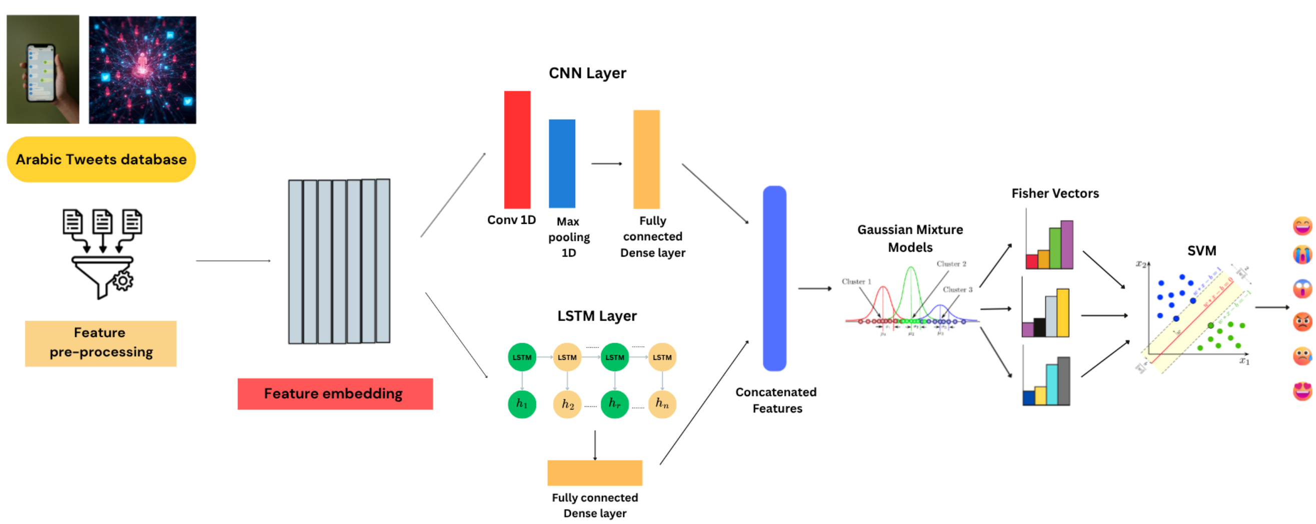

Fig. 2. The Proposed Hybrid Model for Emotions Detection.

and the convolutional layer is then used to illustrate learning from sliding *w*-grams. Convolutional layers build features by filtering the incoming data and combining several filters to produce outputs. CNNs are frequently employed in Natural Language Processing (NLP) and classification tasks. In our proposed model, the CNN consists of a single layer. After feature extraction, the CNN will obtain an embedding layer of size (15098, 300). The CNN will use the training vector as input to predict the data's emotions.

- **Gated Recurrent Unit (GRU):** helps to resolve the vanishing gradient issue in a typical Recurrent Neural Network (RNN). RNNs can remember and apply prior knowledge to source states in the input sequence, which distinguishes them from traditional neural networks. In our suggested model, the GRU is a single-layer network. After feature extraction, the GRU model will receive an embedding layer of size (15098, 300) as input to predict emotions from the data.

- **Bidirectional Gated Recurrent Unit (Bi-GRU):** A bidirectional GRU is a paradigm for sequence processing that consists of two GRUs cooperating. One provides forward-pointing feedback, while the other provides backward-pointing feedback. This bidirectional recurrent neural network only utilizes the input and output gates In our proposed approach, the Bi-GRU model is a single layer.
- **Long Short-Term Memory (LSTM):** This model integrates specialized processes that enable information retention across extended sequences. LSTMs exhibit a chain-like architecture and employ many gates to meticulously regulate which information passes through each node. In our proposed model, an LSTM is used in a single-layer form, taking the embedding layer of size (15098, 300) as input to predict emotions.
- **Bidirectional Long Short-Term Memory (Bi-LSTM):** is an enhancement of the standard LSTM developed in 1997. These represent an extension of LSTM models, in which two LSTMs are applied to the input data. An LSTM is applied to the input sequence in the initial phase (i.e., forward layer). In the subsequent iteration, the LSTM model receives the reverse of the input sequence (i.e., backward layer). Using the LSTM twice effectively captures long-term dependencies, thereby improving the model's accuracy. In our proposed model, the initial layer was the encoding layer, followed by a bidirectional LSTM layer. Subsequently, three dense layers were incorporated to scale, rotate, and transform the vectors, while the softmax activation function was employed to predict the emotions.

The input to our system is the emotion dataset described above, which is fed into the word embedding layer (GloVe). After obtaining the embedding vector, it is utilized as input to both deep learning algorithms, specifically CNN and LSTM. The final layer has been removed from the CNN and LSTM models, allowing them to function as encoders [16]. Additionally, both encoders will produce a latent vector corresponding to the specified input embedding vector.

2) SVM with Generative Kernels: The outputs from the two selected best-performing deep learning models will serve as a combined flattened layer and inputs for the unsupervised models to create generative kernels. In other words, we are developing Support Vector Machine (SVM) kernels based on finite Gaussian Mixture Models (GMM) [17] and infinite mixture models, specifically the Dirichlet Process Gaussian Mixture Model (DPGMM) [18]. These approaches aim to address practical limitations of traditional kernels, such as sensitivity to parameter settings that require careful tuning, the inability to estimate class probabilities, and the risk of overfitting.

We assume $\mathscr{X}$ and $\mathscr{X}'$ two sequences of concatenated feature vectors $\mathscr{X} = \{X_1, \ldots, X_N\}$ and $\mathscr{X}' = \{X'_1, \ldots, X'_{N'}\}$, respectively. These two features will be fitted to the Gaussian Mixture Model (GMM) probability function defined as follows:

$$p(\mathscr{X}) = \sum_{k=1}^{K} w_k \mathscr{N}(\mathscr{X} | \mu_k, \Sigma_k) \tag{1}$$

where $\mathscr{N}(\mathscr{X} | \mu_k, \Sigma_k)$ is the k-th Gaussian distribution with parameters mu_k and Σ_k).

The Fisher kernel is based on extracting Fisher scores $U_{\mathscr{X}}(\Theta) = \bigtriangledown \log(p(\mathscr{X}|\Theta))$ from the generative model and converting them into a kernel to feed SVMs. Each component of $U_{\mathscr{X}}(\Theta)$ is the derivative of the log-likelihood of the sequence $\mathscr{X}$ for a particular parameter of the Gaussian mixture model. The parameters of the Gaussian Mixture Model (GMM) are estimated using two methods. The first is a finite-approach that employs the Expectation-Maximization (EM) algorithm, in which the parameters are optimized to maximize the log-likelihood. The second method is an infinite approach that uses the Dirichlet process (DP), a stochastic process that extends the standard GMM by allowing an unbounded number of clusters. After estimating the GMM parameters, we determine the kernel defined as follows:

$$\mathscr{K}(\mathscr{X}, \mathscr{X}') = U^{tr}_{\mathscr{X}}(\Theta) F^{-1}(\Theta) U_{\mathscr{X}'}(\Theta') \tag{2}$$

where $F^{-1}(\Theta)$ is the Fisher information matrix. Despite the quadratic complexity of the Fisher kernel, it is a highly principled statistical technique. Finally, the GMM Fisher vectors will be used to train an SVM classifier that predicts emotions. Please note that we chose the CNN, LSTM, and Bi-LSTM because they are the two best-performing models when used individually, as shown in the next section.

IV. Results and Discussion

In the first set of our experiments, we tested five different deep learning models, namely CNN, GRU, Bi-GRU, LSM, and Bi-LSTM, on the Arabic emotion Twitter dataset [12]. For feature extraction, we used the most common word-embedding techniques (i.e., word2vec, ELMO, FastText, and GloVe). Table I presents the results of the tested models, using overall accuracy (in percentage), precision, recall, and F1-score as performance metrics. As shown in Table I, in general, CNN, LSTM, and Bi-LSTM performed the best among the tested models. However, these models performed differently using different embedding techniques. For example, although the Bi-LSTM performs best across all tested models when using word2vec embeddings, with an accuracy of 43.16%, its performance improves with ELMO and FastText, reaching 49.9% and 60.74%, respectively. In addition, we noticed that using the GloVe embedding yielded the best performance across almost all the tested models, achieving an overall accuracy of 52%-61%.

For testing the proposed hybrid model, we selected the two best-performing models across all metrics. Thus, CNN, LSTM, and Bi-LSTM with GloVe embedding were chosen due to their superior performance compared to other tested

TABLE I

THE PERFORMANCE OF DEEP LEARNING MODELS USING DIFFERENT EMBEDDING TECHNIQUES.

Embedding	Model	Accuracy (%)	Precision	Recall	F score
Word2vec Embedding	CNN	40.68	0.3932	0.4057	0.3657
	GRU	42.57	0.4149	0.4239	0.3950
	Bi-GRU	40.48	0.3995	0.4033	0.3762
	LSTM	42.72	0.4174	0.4243	0.3996
	Bi-LSM	43.16	0.4198	0.4281	0.3950
Elmo Embedding	CNN	44.61	0.4451	0.4484	0.4120
	GRU	33.01	0.5863	0.3318	0.4582
	Bi-GRU	49.57	0.4933	0.4927	0.4818
	LSTM	47.88	0.4827	0.4846	0.4659
	Bi-LSM	49.97	0.4725	0.5020	0.4654
Fasttext Embedding	CNN	60.00	0.6238	0.6025	0.6088
	GRU	60.45	0.6147	0.6081	0.6059
	Bi-GRU	57.62	0.5996	0.5821	0.5816
	LSTM	61.15	0.6143	0.6123	0.6032
	Bi-LSM	60.74	0.6213	0.6216	0.6151
Glove Embedding	CNN	60.80	0.6192	0.6141	0.6167
	GRU	60.20	0.6109	0.6053	0.5960
	Bi-GRU	59.35	0.5937	0.5988	0.5986
	LSTM	61.74	0.6174	0.6207	0.6190
	Bi-LSM	61.40	0.6165	0.6150	0.6145

models. Next, the flattened layer of the CNN is concatenated with the flattened layer of LSTM or Bi-LSTM and used as prior knowledge to fit GMM or DPGMM. After that, the Fisher score is extracted from the generative model and converted into a kernel to feed the SVM for emotion prediction. The results of our proposed hybrid model for the Arabic emotion dataset are presented in Table II.

As we can see, LSTM+CNN and Bi-LSTM+CNN with GMM Fisher kernels perform similarly, with a slight difference in accuracy, precision, recall, and F-score. Referring to the previous results (in Table I), we observed that our proposed hybrid model improved the performance of both LSTM and Bi-LSTM, increasing accuracy by approximately $24-25\%$ compared to when they were used as standalone models. That is, the accuracy of the two models improved from 61.74% to 85.25% for LSTM, and from 61.40% to 85.21% for Bi-LSTM using the GMM Fisher kernels in SVM. In addition, performance has been further enhanced with the DPGMM Fisher kernels, achieving accuracies of 86.05% and 86.01% for LSTM+CNN and Bi-LSTM+CNN, respectively. This superior performance can be explained by the primary concept of the Fisher kernel, which uses the geometric structure of the statistical manifold by transforming an individual sequence of vectors into a single feature vector defined in the gradient log-likelihood space.

By closely examining the confusion matrix in Figure 3, we see that the proposed LSTM+CNN model with a DPGMM Fisher kernel in SVM distinguishes all classes very well. The easiest classes to distinguish are Fear and Sympathy, with class accuracies of 96.9% and 92.7%, respectively. However, the tweets in the Sadness class can easily be confused with those in the Anger class, as only 77.20% of tweets in this class were correctly classified. Moreover, the Surprise class

TABLE II

THE PERFORMANCE OF THE PROPOSED HYBRID MODEL.

Model	Acc	Presion	Recall	F score
LSTM+ CNN GMM Fisher	85.25	85.38	85.38	85.35
Bi-LSTM+ CNN GMM Fisher	85.21	85.37	85.37	85.30
LSTM+CNN DPGMM Fisher	86.05	86.27	86.19	86.18
Bi-LSTM+CNN DPGMM Fisher	86.01	86.20	86.14	86.10

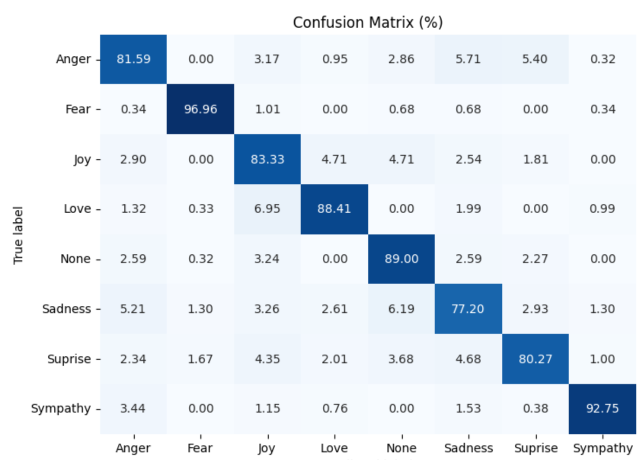

Fig. 3. The Confusion matrix for LSTM+CNN with DPGMM Fisher.

accuracy is 80.27%, which is slightly worse than that of other classes, where the Surprise tweets were misclassified mostly as Sadness or Joy, given that sometimes the language of surprise can express extreme negativity or positivity.

Furthermore, we have compared the performance of our proposed model with recent works published for emotion detection in Arabic text. As shown in Table III, our best

TABLE III

Closely Related Works and Their Performance (Overall Accuracy).

Method	Accuracy
Complement Naïve Bayes algorithm [12]	68.12%
Support vector machine (SVM), and logistic regression [13]	73.39%
Support Vector Machine (SVM) with BoW weighting schema [14]	82.43%
Fully connected neural network structure [15]	59.70%
Combines AraBERT and an attention-based LSTM-BiLSTM model [19]	53.82%
Ensemble method based on MARBERT, Bi-LSTM, Bi-GRU [4]	54.00%
Hybrid deep learning techniques [7]	80.11%
Proposed Method 1: LSTM+ CNN with GMM Fisher kernel	**85.25**%
Proposed Method 2: LSTM+ CNN with DPGMM Fisher kernel	**86.05**%

results using the proposed hybrid model, LSTM+CNN with GMM Fisher kernels and LSTM+CNN with DPGMM Fisher kernels, outperform other approaches reported in the literature. The highest previously reported accuracy for Arabic emotion detection is around 82% [14], while our proposed model achieves an accuracy of up to 86%. It is worth noting that the accuracy reported by the closely related work that uses a hybrid deep learning-SVM model with traditional kernels [7] is only 80.11%. In other words, using the Fisher kernel based on generative models has improved results by around 6% compared to a similar hybrid approach.

V. Conclusion

In this paper, we proposed a hybrid model for detecting emotions in Arabic text. The proposed approach combines CNNs and LSTMs with SVMs, using GMM Fisher kernels for supervised learning. The present work is motivated by the observation that hybrid systems usually perform better than standalone approaches. The proposed approach has been evaluated on the balanced Arabic emotion Twitter dataset. First, we tested several state-of-the-art deep learning models and found that CNN, LSTM, and Bi-LSTM with GloVe embedding give the highest accuracies of 60.80%, 61.74%, and 61.40%, respectively. Then, the best deep learning models were combined with SVM using GMM and DPGMM Fisher kernels. The proposed hybrid model achieves an accuracy of around 85% with the GMM Fisher kernel in SVM and up to 86% with the DPGMM Fisher kernel. The model presented in this paper would also apply to many other problems involving textual data, such as sentiment analysis, fake news detection, and hate speech detection. Potential future work can focus on developing hybrid models using transformers, deep-structured generative models, or ensemble learning.

References

[1] M. Al-A'abed and M. Al-Ayyoub, "A lexicon-based approach for emotion analysis of arabic social media content," in *The International Computer Sciences and Informatics Conference (ICSIC)*, pp. 343–351, 2016.

[2] J. Deng and F. Ren, "A survey of textual emotion recognition and its challenges," *IEEE Transactions on Affective Computing*, vol. 14, no. 1, pp. 49–67, 2021.

[3] A. Thiab, L. Alawneh, and A.-S. Mohammad, "Contextual emotion detection using ensemble deep learning," *Computer Speech & Language*, vol. 86, p. 101604, 2024.

[4] A. Mansy, S. Rady, and T. Gharib, "An ensemble deep learning approach for emotion detection in arabic tweets," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 4, 2022.

[5] R. K. Qarout, J. Y. Samkari, R. M. ALFudhayl, R. H. ALSulami, S. M. Basudan, G. K. AlJuhani, and N. Zamzami, "Emotion detection and classification of different saudi dialects," in *2025 International Symposium on Multimedia (ISM)*, pp. 340–345, IEEE, 2025.

[6] A. Al Maruf, F. Khanam, M. M. Haque, Z. M. Jiyad, F. Mridha, and Z. Aung, "Challenges and opportunities of text-based emotion detection: A survey," *IEEE Access*, 2024.

[7] S. K. Bharti, S. Varadhaganapathy, R. K. Gupta, P. K. Shukla, M. Bouye, S. K. Hingaa, and A. Mahmoud, "Text-based emotion recognition using deep learning approach," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, p. 2645381, 2022.

[8] A. Shmilovici, "Support vector machines," *Data mining and knowledge discovery handbook*, pp. 231–247, 2010.

[9] J. Deng, X. Xu, Z. Zhang, S. Frühholz, D. Grandjean, and B. Schuller, "Fisher kernels on phase-based features for speech emotion recognition," *Dialogues with social robots: Enablements, analyses, and evaluation*, pp. 195–203, 2017.

[10] N. Zamzami and N. Bouguila, "Hybrid generative discriminative approaches based on multinomial scaled dirichlet mixture models," *Applied Intelligence*, vol. 49, no. 11, pp. 3783–3800, 2019.

[11] N. Zamzami and N. Bouguila, "Deriving probabilistic svm kernels from exponential family approximations to multivariate distributions for count data," *Mixture Models and Applications*, pp. 125–153, 2020.

[12] A. Al-Khatib and S. R. El-Beltagy, "Emotional tone detection in arabic tweets," in *Computational Linguistics and Intelligent Text Processing: 18th International Conference, CICLing 2017, Budapest, Hungary, April 17–23, 2017, Revised Selected Papers, Part II 18*, pp. 105–114, Springer, 2018.

[13] A. A. AlFutamani and H. H. Al-Baity, "Emotional analysis of arabic saudi dialect tweets using a supervised learning approach.," *Intelligent Automation & Soft Computing*, vol. 29, no. 1, 2021.

[14] M. A. Al-Hagery, M. A. Al-Assaf, and F. M. Al-kharboush, "Exploration of the best performance method of emotions classification for arabic tweets," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 2, pp. 1010–1020, 2020.

[15] M. Abdullah and S. Shaikh, "Teamuncc at semeval-2018 task 1: Emotion detection in english and arabic tweets using deep learning," in *Proceedings of the 12th international workshop on semantic evaluation*, pp. 350–357, 2018.

[16] E. M. Onyema, P. K. Shukla, S. Dalal, M. N. Mathur, M. Zakariah, and B. Tiwari, "Enhancement of patient facial recognition through deep learning algorithm: Convnet," *Journal of Healthcare Engineering*, vol. 2021, no. 1, p. 5196000, 2021.

[17] B. Karatay, D. Beştepe, K. Sailunaz, T. Özyer, and R. Alhajj, "CNN-Transformer based emotion classification from facial expressions and body gestures," *Multimedia Tools and Applications*, vol. 83, pp. 23129–23171, Mar. 2024.

[18] S. Lee, W. Chang, J. Mateu, H. Lee, and J. Park, "A spatio-temporal dirichlet process mixture model on linear networks for crime data," 2025.

[19] H. Elfaik *et al.*, "Combining context-aware embeddings and an attentional deep learning model for arabic affect analysis on twitter," *IEEE Access*, vol. 9, pp. 111214–111230, 2021.