Classification of Arabic Social Media Texts Based on a Deep Learning Multi-Tasks Model

Ali A. Jalil
University of AlKafeel, Najf, Iraq

Ahmed H. Aliwy
Department of Computer Science, Faculty of Computer Science and Mathematics, University of Kufa, Najaf, Iraq

Classification of Arabic Social Media Texts Based on a Deep Learning Multi-Tasks Model

Cover Page Footnote

Classification of Arabic Social Media Texts Based on a Deep Learning Multi-Tasks Model

Ali A. Jalil [a,b,*], Ahmed H. Aliwy [b]

[a] University of AlKafeel, Najf, Iraq
[b] Department of Computer Science, Faculty of Computer Science and Mathematics, University of Kufa, Najaf, Iraq

Abstract

The proliferation of social networking sites and their user base has led to an exponential increase in the amount of data generated on a daily basis. Textual content is one type of data that is commonly found on these platforms, and it has been shown to have a significant impact on decision-making processes at the individual, group, and national levels. One of the most important and largest part of this data are the texts that express human intentions, feelings and condition. Understanding these texts is one of the biggest challenges that facing data analysis. It is the backbone for understanding people, their orientations, and making decisions in many cases and thus predicting their behavior. In this paper, a model was proposed for understanding texts that written by people on social media platforms, and hence knowing people's attitudes within specific topics, the emotion of those people, positivity, negativity, and neutrality. Also, it extracts emotion of those people. In this context, the system solves many tasks in natural language processing therefore it uses many techniques including topic classifier, sentiment analyzer, sarcasm detector and emotion classifier. CNN-BiLSTM was used for topic classifier, sentiment analyzer, sarcasm detector, and emotion classifier where (f-measure, accuracy) were (97,97.58) %, (84,86) %, (95,97) %, and (82,81.6) % respectively.

Keywords: Social media, Text classification, Sarcasm, Emotion, Sentiment, Topic, Deep learning, Algorithms, Evaluation metrics, Datasets

1. Introduction

The internet has evolved significantly since its inception as a tool for researchers, with the emergence of the World Wide Web and social media platforms. Social media has become an integral part of daily life, transforming communication, information sharing, and business operations. Globally, there are approximately 4.9 billion social media users, with the Middle East and North Africa region having over 200 million users. Understanding social media content is crucial as it can impact political decisions, market trends, and more. Social media platforms, such as Twitter, have become prominent channels for communication and expression in the digital era. With its unique features, including a limit of 280 characters per tweet, Twitter enables users to share their thoughts and ideas in concise and explicit messages. As one of the most widely used social media platforms, Twitter offers a direct connection with the masses, making it a valuable source of data for understanding user opinions, sentiments, emotions, and topics. However, the brevity and informal nature of Twitter content pose challenges for accurately analyzing and interpreting the nuanced aspects of social media content, such as sarcasm, emotion, sentiment, and topic. This article, highlight the importance of text classification in understanding social media content, with a specific focus on Twitter, and discuss the challenges, opportunities, evaluation metrics, and datasets associated with text classification in this context. Sentiment analysis, also known as opinion mining, is the task of determining the sentiment or emotion expressed in a piece of text [1−3]. The concept of sentiment was initially introduced by Das and Chen [4], as well as Tong [5], who examined the sentiment of the market through automated analysis of text. Some of the first

researchers to explore sentiment analysis and its relationship with Natural Language Processing techniques were Turney [6], Pang and their colleagues [7], as well as Nasukawa and Yi [8], as discussed in their respective publications. This is a crucial task in social media analysis as it allows researchers to understand public opinion on a particular topic or product [9]. Several studies have applied machine learning and deep learning techniques to sentiment analysis of social media data, with varying levels of success. Researchers have applied sentiment analysis to a wide range of social media platforms, including Twitter, Facebook, and Instagram.

Li et al. (2020) proposed a bidirectional LSTM model with a self-attention mechanism and multi-channel features (SAMF-BiLSTM). The method models existing linguistic knowledge and sentiment resources in sentiment analysis tasks to form different feature channels, and utilizes a self-attention mechanism to enhance sentiment information [10].

Sangar et al. (2020), addressed the challenge of adapting sentiment lexicon to new domains of the same genre by proposing an unsupervised transfer learning approach that creates new learning insights for multiple domains. An incremental learning methodology is used to learn polarity seed words from automatically selected source domains, which is then transferred to the target domains for sentiment lexicon generation. The process uses latent semantic analysis technique and unlabeled training data from both source and target domains. The experiment was performed on 24 domains of the same genre, consumer product reviews, and the proposed model achieved the best results compared to competitive baselines with a maximum accuracy of 86% [11].

El Karfi et al. (2022) has established the effectiveness of attentions-based models in Arabic sentiment analysis. This work experiment with the use of two transformer-based models, AraBERT and CAMeLBERT, as well as an ensemble model, in order to attain optimal performance [12]. Benroubal et al. (2022), were proposed approach involves using sentiment analysis to filter social media content that may have a negative emotional impact on users. They intend to utilize the Twitter API to gather posts, and a natural language processing tool to classify the emotions expressed in the content into five basic categories. They clarify that will help mitigate the negative effects of certain posts on platforms like Facebook and Twitter [13]. Qin, Y. et al. (2023), a new algorithm called TCN-BiGRU-DATT was proposed. The algorithm uses ALBERT to obtain a vector representation of the text, followed by TCN and BiGRU networks to extract emotional information through dual pathway feature extraction.

The dual attention mechanism then allocates global weight to key information in the semantic features and the emotional features are spliced and fused. Finally, a Softmax classifier is applied for emotion classification. The proposed algorithm achieved accuracy, recall, and F1 value of 92.33%, 91.78%, and 91.52%, respectively [14]. Topic modeling is another text mining technique that has been widely used to understand social media content. Topic modeling is a method of uncovering latent topics in a corpus of text data [15]. This can be used to identify trending topics on social media, as well as to understand the spread of information and misinformation. Albalawi R et al., (2020) investigated the topic modeling subject and its common areas of application, methods, and tools. An examination and comparison of five frequently used topic modeling methods are conducted to demonstrate their practical benefits in detecting important topics when applied to short textual social data. It was found that the latent Dirlet allocation and non-negative matrix factorization methods delivered the most meaningful extracted topics and had good results [16]. Teh PA et al., (2022) utilized a set of topic modeling to help policymakers and environmental communities understand public opinions on plastic pollution issues through the analysis of social media data. Five topic modeling techniques were applied to the data to identify popular topics of online conversations. The experimental results indicate that some of these topic modeling techniques are effective in identifying crucial topics related to plastic pollution [17]. Mihunov V et al., (2022) presented a framework that can provide disaster impact information sourced from social media in a timely manner. The framework was tested using Hurricane Harvey as a well-studied and data-rich case. The raw Twitter data was filtered based on keywords, location, and tweet attributes and then the latent Dirichlet allocation was applied to categorize the tweets into topics useful to emergency managers. Significant correlations were found between the nine relevant topics and population density, but not flood depth and damage, indicating the need for further research into the suitability of social media data for disaster impact modeling [18]. This study proposes deep learning-based model to classify Arabic social media content in term of (sarcasm, emotion, sentiment and topic).

2. Materials and methods

2.1. Text mining, text classification and social media content

Text mining, is the process of using analytical techniques to extract information from a large collection of text data. By using algorithms, one can

analyze large amounts of text and identify key themes and relevant information [19].

Text classification, also known as text categorization or text tagging, is a task in natural language processing (NLP) that involves automatically assigning predefined categories or labels to a given text based on its content [20]. Text classification has various applications, such as spam detection, sentiment analysis, topic categorization, and intent recognition in chatbots. Text classification for social media content poses several challenges due to the unique characteristics of social media data. Firstly, the brevity and informal nature of Twitter content make it challenging to accurately capture the nuances of sarcasm, emotion, sentiment, and topic. Sarcasm, in particular, can be expressed implicitly through tone, context, and word choice, making it difficult to detect without proper contextual understanding. Emotions and sentiments can be expressed in abbreviated forms or with the use of emojis and hashtags, which require careful consideration in the text classification process. Additionally, the dynamic nature of social media data, with constantly evolving trends and discussions, makes it challenging to develop robust and generalizable text classification models. Another challenge is the availability of high-quality labeled datasets for training and evaluation, as social media data can be noisy, biased, and subjective. Ensuring the reliability and representativeness of datasets is crucial for achieving accurate and generalizable text classification models [21].

2.2. The proposed model

Social media platforms have become a rich source of information with a vast amount of user-generated content. Extracting relevant information from social media texts can provide valuable insights into various domains, including politics, religion, sports, literature, technology, arts, and economics. However, understanding the content of social media posts is challenging due to the informal and unstructured nature of text in social media. To address this challenge, it is essential to develop a specialized system for text understanding that is tailored to the specific purpose for which the text needs to be understood. This requires determining the desired information to be extracted from the text, including orientations within specific topics, sentiment analysis, and emotional analysis. This research, propose a hybrid model based on Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) to achieve these objectives. The proposed hybrid model combines the strengths of CNN and LSTM to effectively extract information from social media texts. CNN is known for its ability to capture local features and patterns in data, making it suitable for tasks that require spatial information analysis, such as image processing. LSTM, on the other hand, is a type of recurrent neural network that is capable of modeling long-term dependencies and capturing sequential patterns, making it well-suited for tasks that involve analyzing sequences of data, such as text processing. The blueprint of the proposed hybrid model is illustrated in Fig. 1. And details of it are described as bellow.

2.3. Data collecting and acquisition

A dataset of 9 million Arabic tweets was constructed through the utilization of the Twitter API, which collected data based on predefined words that corresponded to 9 topics: Art, Literature, Politics, Sport, Economy, Religion, Science, Health and Tech. Each tweet was subsequently categorized into its respective topic. Fig. 2 shows the distributions of topics in topics dataset. In addition, three types of datasets for sarcasm, sentiment and emotion where used.

Almahdawi et al. (2019) [22] collected a dataset of Social Media posts in the Iraqi dialect that is annotated according to Ekman's six basic emotions, including Anger, Disgust, Fear, Happiness, Sadness, and Surprise. This dataset, containing 1353 rows, is specifically designed for analyzing emotions in Arabic text. Fig. 3 show the distribution of the six classes in the emotion dataset.

For sarcasm detection, Abu Farha et al., in 2021 published the ArSarcasm-v2 dataset [23], which includes 15,548 Arabic tweets labeled as sarcasm and non-sarcasm.

For sentiment analysis, four datasets were used: Sentiment Analysis Dataset - SS2030 [24], An Arabic Speech-Act, Sentiment Corpus of Tweets (ArSAS) [25], Twitter Data set for Arabic Sentiment Analysis [26], and ArSarcasm-v2 [23]. These datasets consisted of 4214, 21,000, 348,797, and 15,548 tweets, respectively. They were merged to train the sentiment analysis classifier, and the tweets were distributed into three categories: Neutral (NEU), Positive (POS), and Negative (NEG), as shown in Fig. 4. Due to the insufficiency of gathered data for sarcasm and emotion dataset, an augmentation process is necessary to effectively train a deep learning model. To achieve this, the nlpaug library is utilized alongside a sequential model consisting of a BERT model with a pretrained model proposed by G. Inoue et la. [27], fast text with Skip-gram Arabic version of pretrained model proposed by P. Bojanowski et la. [28], glove with pre-trained model
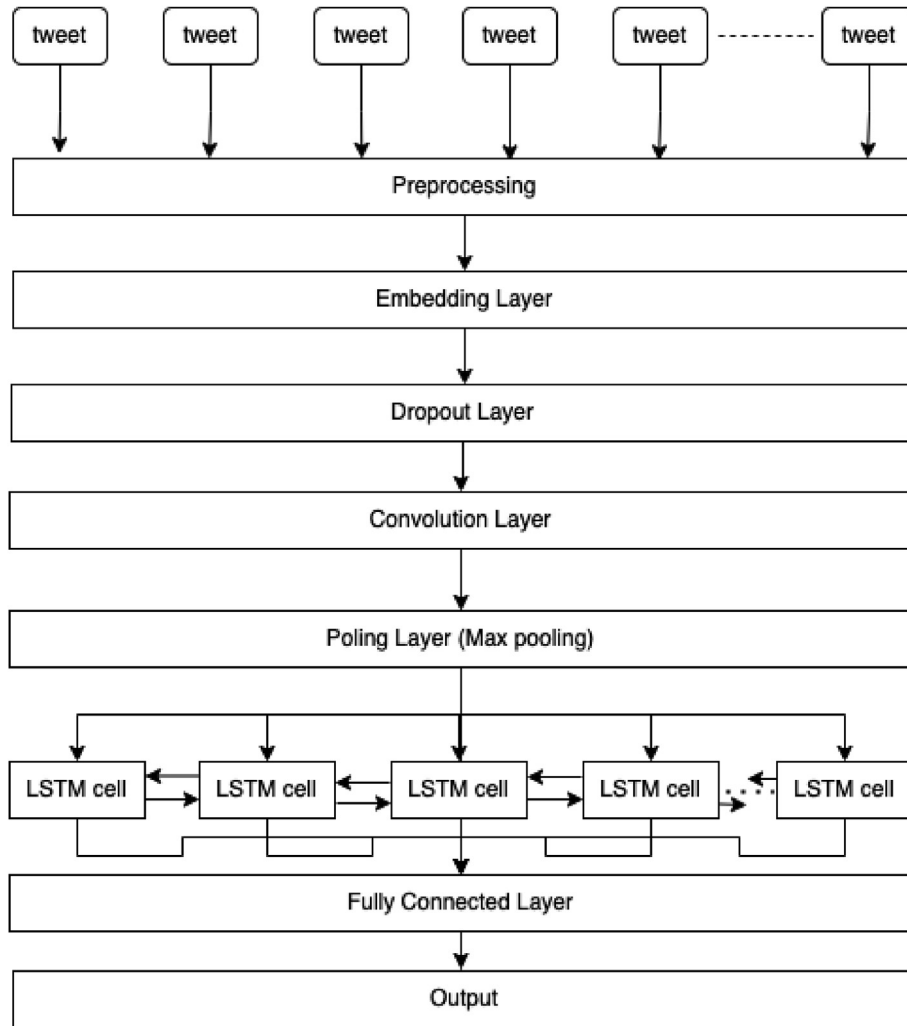
Fig. 1. The proposed model to classify Arabic social media textual contents.
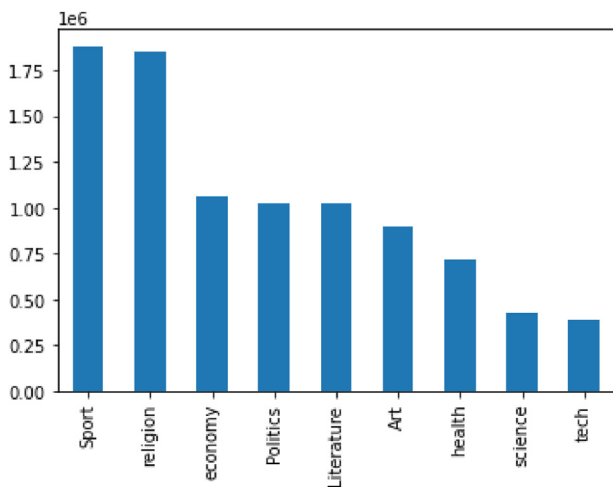

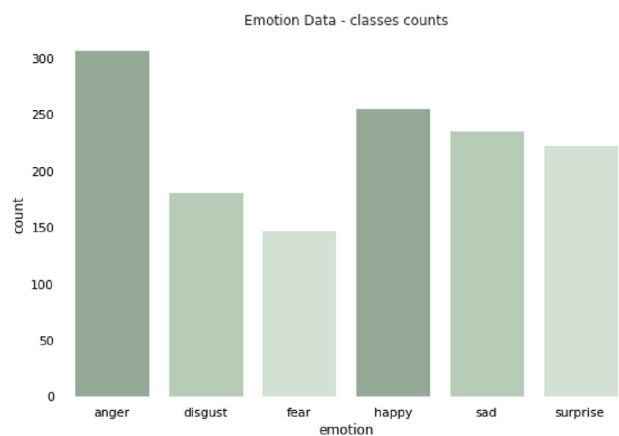*Fig. 2. Distributions of topics in topics dataset.*

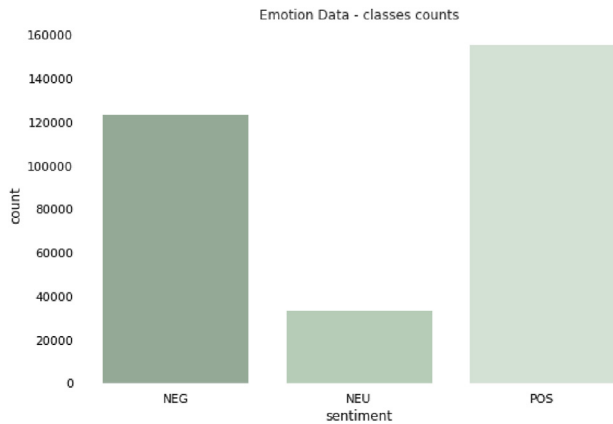*Fig. 3. Distribution of the six classes in the emotion dataset.*

Fig. 4. Distribution of sentiment's classes in the sentiment dataset.

proposed by Arabic language technology group [29], and word2vec with pretrained model proposed by A. B. Soliman et la [30].

2.4. Preprocessing

The collected data, from Twitter, have many unrelated data, so it is preprocessed to be reliable and can be used for many applications and tasks. Since the collected tweets are in Arabic language so special processing tools should be used.

2.4.1. Remove unnecessaries

The collected data has unnecessaries symbols and words that should be removed in the first step of processing. Samples of these unnecessary symbols/words are:

- Emoji: A symbol or icon used to express a situation or an emotion.
- User handler: a word or name that comes after the (@) sign which represents a Twitter account.
- Html tags: special characters used in a hypertext markup language.
- Email address.
- Diacritics.

2.4.2. Normalization

The Arabic language consists of 28 basic letters in addition to the hamza. Some of these letters take different forms depending on their position in speech.

For example, the letter (alif) takes the following forms (إ,آ,أ,ا) or the literal alif (ى), as well as the overlap between the letters Haa (ه) and Taa marbota (ة) in addition to the hamza on one of the basic letters such as (ئ,ؤ) so these letters must be unified in one form by the normalization process [31].

2.4.3. Tokenization

Two types of tokenization can be done for Arabic language. The first one is by extracting the word as a token from the text based on spaces and punctuations while the second one is by extracting sub word as a token. These two types are shown in Fig. (5) [32]. An example of second type tokenizer is Byte pair encoding (BPE).

2.4.4. Lemmatization

It is the process of using the morphological analysis of the word and removing the inflectional additions of the word to return it to its original form based on semantic of this word to produce a lemma. i.e. some variants of words can be based on one lemma [7] for example the lemma for the word "كاتبة" and "كاتب" is "كاتب".

2.4.5. Stop word removal

They are the most common words in speech (such as the names of months, pronouns, prepositions and connections, numbers and digits, and others) whose presence may affect the performance, so they must be filtered before starting the treatment process to obtain better results when treating. Stop word removal is application dependent that mean the list of stop words is different between two applications. For example, in POS tagging system the prepositions are very important while in IR system they should be removed. In this work, 481 Arabic lemmas are used as stop words.

2.5. Classifier structure

A classifier is constructed with the following layers to categorize Arabic text based on sarcasm, emotion, sentiment, and topic.

2.5.1. Word embedding

The first layer of proposed model is a word embeddings layer, A word embedding layer in a deep

Sentence		سنحبهم جميعا مع أبنائهم
Tokenization	A	سنحبهم + جميعا + مع + أبنائهم
	B	س + نحب + هم + جميعا + مع + أبناء + هم

Fig. 5. Distribution of sentiment's classes in the sentiment dataset.

learning model is a type of layer that maps each word in the input text to a high-dimensional vector space. The goal of a word embedding layer is to capture the semantic relationships between words and represent them in a way that can be easily used by a neural network [33]. A word embedding layer typically takes a sequence of words as input, such as a sentence or a paragraph, and generates a matrix of embeddings as output, where each row of the matrix represents the embedding for a single word in the sequence. The size of the embedding vector is usually much smaller than the vocabulary size, which can be several thousand or even millions of words [34]. The word2vec word embedding techniques with skip gram of version of pretrained model proposed by A. B. Soliman et la [30]. with vector length equal to 100 were used as word embeddings layer in proposed model.

2.5.2. Dropout layer

Dropout is a regularization technique used in deep learning neural networks to prevent overfitting. It works by randomly dropping out (i.e., setting to zero) a certain percentage of the neurons in the layer during each training iteration. This has the effect of making the network more robust and less likely to rely on any one particular set of neurons. During training, the dropout layer randomly selects a subset of neurons to be dropped out based on a probability value which proposed model is set to 0.25. The remaining neurons are then scaled up by a factor of 1/(1 - p), where p is the probability of dropout, to ensure that the total input to the next layer remains constant.

2.5.3. CNN layer

The convolutional layer is responsible for extracting features from the input text. The input to the convolutional layer is a word embedding, where each row represents a word vector. The convolutional layer applies a set of filters with a fixed size to the input matrix, moving them over the rows of the matrix to produce a feature map. The output of the convolutional layer is a set of feature maps, where each map corresponds to a specific filter. To introduce non-linearity and increase the model's expressive power, the output of the convolutional layer is passed through the Rectified Linear Unit (ReLU) activation function [35]. Word embeddings are typically represented as vectors of real numbers, and it is possible for some of these values to be negative. The ReLU function returns the input value if it's positive, and zero otherwise. After the activation function, the output is passed through a pooling layer, which reduces the dimensionality of the feature maps by down sampling them. The used pooling layer is max pooling, which takes the maximum value across a fixed window of the feature map. The output of the pooling layer is passed to the next stage.

2.5.4. BiLSTM layer

Bidirectional Long Short-Term Memory (BiLSTM) is a type of recurrent neural network (RNN), unlike traditional RNNs that process input sequences in a single direction, BiLSTMs process input sequences in both forward and backward directions simultaneously. This allows the network to capture not only the preceding context but also the succeeding context for each word in the input sequence. The input to a BiLSTM layer is the features captured by previous layer. The BiLSTM layer processes the input in both forward and backward directions, using two sets of hidden states and cell states. The output of the BiLSTM layer is a matrix of concatenated hidden states, which captures both the forward and backward context for each word in the input sequence. The output of the BiLSTM layer then passed through fully connected layers, followed by a softmax [35] layer to produce the final classification output. BiLSTMs are particularly useful for text classification tasks where context is important, such as sentiment analysis or named entity recognition. By capturing both the preceding and succeeding context for each word, BiLSTMs can better capture the semantics and meaning of the input text.

2.6. Evaluation metrics

There are many evaluation metrics can be used for validity of any proposed algorithm, method, system or techniques. The most well-known evaluation metrics for classification are Precision, Recall, F measure, and accuracy. Precision indicates the percentage of correct prediction of the class, while the recall indicates the percentage of true class which retrieves, whereas the F measure represent the summary of precision and recall and can calculated by following formulas:

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$F = \frac{2PR}{P + R} \tag{3}$$

Table 1. *Summary of results obtained from applying CNN-BiLST model with different datasets.*

Data Set	Epochs	CNN, Filters, Kernal	Pool Size	LSTM Units	Batches	P	R	F1	Accuracy
Sarcasm	34	64,3	2	32	64	97%	94%	95%	97%
Sentiment	16	64,3	2	32	128	86%	83%	84%	86%
Emotion	37	64,3	2	70	64	82%	82%	82%	81.6%
Topics	14	32,3	2	100	256	97%	97%	97%	97.58%

$$Accuracy = \frac{TP + TN}{N} \qquad (4)$$

Where: TP, FP, FN, N are true positive, false positive, false negative and Count of classified texts respectively.

3. Results and discussion

This research involved developing hybrid deep learning (CNN-BiLSTM) model for both multi-classification tasks, including sentiment, emotion, and topic, as well as binary classification tasks, specifically sarcasm classification. The datasets used for training and building these models were described earlier, and the Hold-out cross-validation method was utilized with an 80% training and 20% testing split, optimizer ADAM with learning rate of 0.0001, dropout of 0.25 and loos function Categorical Cross Entropy (CCE) for emotion, sentiment and topic detection while a Binary Cross Entropy (BCE) for sarcasm detection. The machine used for the experiments is equipped with 12th Generation Intel® Core™ i7 Processors, 32 GB RAM, a 1 TB SSD hard drive, and a GeForce RTX 3050 GPU.

The sarcasm detection model proposed in this study was trained using an augmented sarcasm dataset, resulting in an impressive accuracy of 97%. The model also demonstrated high macro precision of 97%, macro recall of 94%, and an f1-score of 95%. The hyperparameters utilized in the model included a CNN filter count of 64, kernel size of 3, pool size of 2, LSTM unit count of 32, batch size of 64, and 33

Table 2. *The Proposed model in compare with the past works.*

Method	Task	Metric	
		Type	Value
hybrid approach using both corpus based and dictionary based Dheeraj Kumar et al. [36]	Sentiment	Accuracy	84.3%
multinomial Naïve Bayes Ismail H et al. [37]	Sentiment	Accuracy	81.34%
logistic regression classifier Li et al. [38]	Sentiment	Accuracy	68.99%
support vector machine Lopez et al. [39]	Emotion	Accuracy	76.36%
Bidirectional LSTM with self-attention mechanism and multi-channel features Li et al. [10]	Sentiment	Accuracy	89.7%
Multidomain Sentiment Lexicon Sangar et al. [11]	Sentiment	Accuracy	86%
AraBERT, CAMeLBERT, Arabic ALBERT, GigaBERT and ensemble model El Karfi et al. [12]	Sentiment	Macro-F1	96%, 94%, 86%, 88%,84%, 89%
TCN-BiGRU and Dual Attention Qin et al. [14]	Sentiment	Macro-F1	92.33%
LDA Valencia et al. [40]	Topic	coherence score	0.4896
LSTM and LDA Sit M et al. [41]	Topic	Macro-F1	73%
Non-Negative Matrix Albalawi R et al. [16]	Topic	Macro-F1	61.6%
LDA, Hierarchical Dirichlet Process, Latent Semantic Indexing Teh PA et al. [17]	Topic	coherence score	0.4964, 0.3503, 0.3440
CNN-BiSTM	**Topic**	**Macro-F1**	**97%**
CNN-BiSTM	**Sarcasm**	**Macro-F1**	**94%**
CNN-BiSTM	**Sentiment**	**Macro-F1**	**84%**
CNN-BiSTM	**Emotion**	**Macro-F1**	**82%**

epochs. For sentiment detection in social media texts, the proposed model was trained using a sentiment dataset, achieving an accuracy of 86%. The model also exhibited a macro precision of 86.9%, macro recall of 83%, and an f1-score of 84%. The hyperparameters employed in this model were a CNN filter count of 64, kernel size of 3, pool size of 2, LSTM unit count of 128, batch size of 128, and 10 epochs. The proposed model for emotion detection in social media texts was trained using an augmented emotions dataset and achieved an accuracy of 81.6%. The model demonstrated a precision of 82%, recall of 82%, and an f1-score of 82%. The hyperparameters used in this model were a CNN filter count of 64, kernel size of 3, pool size of 2, LSTM unit count of 70, batch size of 64, and 45 epochs. Lastly, the proposed model for topic detection in social media texts was trained using a topics dataset and achieved an accuracy of 97.58%. The model also displayed high precision, recall, and f1-score of 97%. The hyperparameters employed in this model were a CNN filter count of 32, kernel size of 3, LSTM unit count of 100, batch size of 256, and 14 epochs.

The model that was suggested could identify multiple categories within the same text, a feature that was not available in any of the previous models (see Table 1). Table 2 provides a comparison between the proposed model and the earlier works.

4. Conclusion

The CNN-BiLSTM model addresses the social media text processing challenges by combining the strengths of the CNN and BiLSTM architectures. The CNN component can capture local patterns in the text, such as word associations, while the BiLSTM component can capture long-term dependencies between words. The model can be trained on different datasets that focus on different aspects of social media data, such as topic, emotion, sarcasm, and sentiment, and can then be used to analyze new social media posts and extract relevant information. The model can also be adapted to new datasets and domains, making it a flexible and scalable solution for social media analysis. The CNN-BiLSTM model that can identify the topic of social media content, detect sarcasm, emotion, and sentiment and has several potential applications. Some examples include its use by companies to better understand how their products or services are being perceived on social media platforms, and by customer service teams to automatically detect the sentiment and emotion of customer feedback on social media platforms, thereby improving customer satisfaction. The

model can also be used by political campaigns to track sentiment and emotion around specific political issues on social media, by news organizations to automatically categorize news stories by topic, sentiment, and emotion, and by researchers to analyze social media data for studies on topics such as public opinion, social trends, and mental health. Overall, a wide range of potential applications in marketing, customer service, politics, news, research, and many other fields where social media content analysis is valuable are provided by this model.

Acknowledgements

References

[1] Giachanou A, Crestani F. Like it or not: a survey of twitter sentiment analysis methods. ACM Comput Surv Jun. 2017; 49(2):1−41. https://doi.org/10.1145/2938640.

[2] Kumar A, Sharma A. Twitter sentiment analysis using machine learning: a review. International Research Journal of Engineering and Technology; 2020 [Online]. Available: www.imdb.com.

[3] Wankhade M, Rao ACS, Kulkarni C. A survey on sentiment analysis methods, applications, and challenges. Artif Intell Rev, Oct 2022;55(7):5731−80. https://doi.org/10.1007/s10462-022-10144-1.

[4] Das S, Chen M. Yahoo! for Amazon: extracting market sentiment from stock message boards. In: Asia pacific finance association annual conf. APFA); 2001.

[5] Tong R. An operational system for detecting and tracking opinions in on-line discussions. In: Working notes of the SIGIR workshop on operational text classification. Louisianna: New Orleans; 2001. p. 1−6.

[6] P. D. Turney, "Thumbs up or thumbs down? Semantic Orientation Applied to Unsupervised Classification of Reviews." [Online]. Available: http://www.google.com.

[7] Pang B, Lee L, Vaithyanathan S. Thumbs up?. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing - EMNLP '02. Morristown, NJ, USA: Association for Computational Linguistics; 2002. p. 79−86. https://doi.org/10.3115/1118693.1118704.

[8] Nasukawa T, Yi J. Sentiment analysis. In: Proceedings of the 2nd international conference on Knowledge capture. New York, NY, USA: ACM; Oct. 2003. p. 70−7. https://doi.org/10.1145/945645.945658.

[9] Wang Y, Guo J, Yuan C, Li B. Sentiment analysis of twitter data. Appl Sci Nov. 2022;12(22):11775. https://doi.org/10.3390/app122211775.

[10] Li W, Qi F, Tang M, Yu Z. Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification. Neurocomputing Apr. 2020;387:63−77. https://doi.org/10.1016/j.neucom.2020.01.006.

[11] Sanagar S, Gupta D. Unsupervised genre-based multidomain sentiment lexicon learning using corpus-generated polarity seed words. IEEE Access 2020;8:118050−71. https://doi.org/10.1109/ACCESS.2020.3005242.

[12] El Karfi I, El Fkihi S. An ensemble of Arabic transformer-based models for Arabic sentiment analysis. Int J Adv Comput Sci Appl 2022;13(8). https://doi.org/10.14569/IJACSA.2022.0130865.

[13] Benrouba F, Boudour R. Emotional sentiment analysis of social media content for mental health safety. 2022. https://doi.org/10.21203/rs.3.rs-2170906/v1.

[14] Qin Y, Shi Y, Hao X, Liu J. Microblog text emotion classification algorithm based on TCN-BiGRU and dual attention. Information 2023;vol. 14(2). https://doi.org/10.3390/info14020090.

[15] Paul MJ, Dredze M. Discovering health topics in social media using topic models. PLoS One 2014;9(8):e103408.

[16] Albalawi R, Yeap TH, Benyoucef M. Using topic modeling methods for short-text data: a comparative analysis. Front Artif Intell 2020;3. https://doi.org/10.3389/frai.2020.00042.

[17] Teh PL, Piao S, Almansour M, Ong HF, Ahad A. Analysis of popular social media topics regarding plastic pollution. Sustainability 2022;14(3):1709. https://doi.org/10.3390/su14031709.

[18] Mihunov VV, Jafari NH, Wang K, Lam NSN, Govender D. Disaster impacts surveillance from social media with topic modeling and feature extraction: case of Hurricane Harvey. International Journal of Disaster Risk Science Oct. 2022;13(5):729–42. https://doi.org/10.1007/s13753-022-00442-1.

[19] He W, Zha S, Li L. Social media competitive analysis and text mining: a case study in the pizza industry. Int J Inf Manag Jun. 2013;33(3):464–72. https://doi.org/10.1016/j.ijinfomgt.2013.01.001.

[20] Panchendrarajan R, Amaresan A. Bidirectional LSTM-CRF for named entity recognition. 2018.

[21] Arif MH, Li J, Iqbal M. Solving social media text classification problems using code fragment-based XCSR. In: IEEE 29th international conference on tools with artificial intelligence (ICTAI). IEEE; 2017. p. 485–92. https://doi.org/10.1109/ICTAI.2017.00080.

[22] Almahdawi AJ, Teahan WJ. A new Arabic dataset for emotion recognition. In: Advances in intelligent systems and computing. Springer Verlag; 2019. p. 200–16. https://doi.org/10.1007/978-3-030-22868-2_16.

[23] Abu Farha I, Zaghouani W, Magdy W. Overview of the WANLP 2021 shared task on sarcasm and sentiment detection in Arabic. In: Proceedings of the sixth Arabic Natural Language Processing workshop, Kyiv, Ukraine (Virtual): Association for Computational Linguistics; Apr. 2021. p. 296–305 [Online]. Available: https://aclanthology.org/2021.wanlp-1.36.

[24] Alyami SN, Olatunji SO. Application of support vector machine for Arabic sentiment classification using twitter-based dataset. J Inf Knowl Manag 2020;19(1):2040018. https://doi.org/10.1142/S0219649220400183.

[25] Elmadany A, Mubarak H, Magdy W. ArSAS: an Arabic speech-act and sentiment corpus of tweets. In: Proceedings of the 3rd workshop on open-source Arabic corpora and processing tools; 2018. p. 13–7. *co-located with LREC 2018.*

[26] Abdulla NA, Mahyoub NA, Shehab M, Al-Ayyoub M. Arabic sentiment analysis: corpus-based and lexicon-based. In: IEEE conference on applied electrical engineering and computing technologies (AEECT); Dec. 2013. p. 1–6. Amman, Jordan.

[27] Inoue G, Alhafni B, Baimukan N, Bouamor H, Habash N. The interplay of variant, size, and task type in Arabic pre-trained language models. In: Proceedings of the sixth Arabic

Natural Language Processing workshop, kyiv, Ukraine (on-line): Association for Computational Linguistics; Apr. 2021.

[28] Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching word vectors with subword information. Jul. 2016.

[29] Arabic Language Technologies Group. Arabic word embeddings model. Qatar Computing Research Institute; 2017.

[30] Soliman AB, Eisa K, El-Beltagy SR. 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.

[31] Darwish K, Magdy W, Mourad A. Language processing for Arabic microblog retrieval. In: Proceedings of the 21st ACM international conference on Information and knowledge management; 2012. p. 2427–30.

[32] M. Alhanjouri, "Pre processing techniques for Arabic documents clustering," Int J Eng Manag Res, no. 7, [Online]. Available: www.ijemr.net.

[33] Schnabel T, Labutov I, Mimno D, Joachims T. Evaluation methods for unsupervised word embeddings. In: Proceedings of the 2015 conference on empirical methods in Natural Language Processing. Stroudsburg, PA, USA: Association for Computational Linguistics; 2015. p. 298–307. https://doi.org/10.18653/v1/D15-1036.

[34] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In: Proceedings of the international conference on learning representations. ICLR); 2013.

[35] Bergstra J, Ca JB, Ca YB. Random search for hyper-parameter optimization yoshua bengio. 2012 [Online]. Available: http://scikit-learn.sourceforge.net.

[36] Kumar A, Sebastian TM. Sentiment analysis on twitter. 2012 [Online]. Available: www.youtube.com.

[37] Ismail HM, Harous S, Belkhouche B. A comparative analysis of machine learning classifiers for twitter sentiment analysis. Research in Computing Science Dec. 2016;110(1):71–83. https://doi.org/10.13053/rcs-110-1-6.

[38] Li M, Ch'ng E, Chong AYL, See S. "Multi-class Twitter sentiment classification with emojis. Ind Manag Data Syst 2018;118(9):1804–1820, Sep. https://doi.org/10.1108/IMDS-12-2017-0582.

[39] López-Chau A, Valle-Cruz D, Sandoval-Almazán R. Sentiment analysis of twitter data through machine learning techniques. 2020. p. 185–209. https://doi.org/10.1007/978-3-030-33624-0_8.

[40] Valencia JD, Al Joseph TL, Centino NM, Fabito BS, Imperial J.M.Rodriguez RL, et al. Understanding anonymous social media posts using topic modeling. In: 2019 IEEE 11th international conference on humanoid, nanotechnology, information technology, Communication and Control, Environment, and Management (HNICEM). IEEE; Nov. 2019. p. 1–4. https://doi.org/10.1109/HNICEM48295.2019.9072791.

[41] Sit MA, Koylu C, Demir I. Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: a case study of Hurricane Irma. Int J Digit Earth, Nov 2019;12(11):1205–29. https://doi.org/10.1080/17538947.2018.1563219.