New model for emotion detecting from French text using bidirectional long short-term memory

Aya Adel[1,2], Shereen A. Taie[1], Esraa Elhariri[1], Mohamed Hasan Ibrahim[3]
[1]Department of Computer Science, Faculty of Computers and Artificial Intelligence, Fayoum University, Fayoum, Egypt
[2]Department of Artificial Intelligence, School of Artificial Intelligence, Egyptian Russian University, Cairo, Egypt
[3]Department of Information System, Faculty of Computers and Artificial Intelligence, Fayoum University, Fayoum, Egypt

Article Info

Article history:

Received May 29, 2024
Revised Oct 15, 2024
Accepted Nov 28, 2024

Keywords:

Bidirectional long short-term memory
Camembert tokenizer
Embedding layer
Emotion detection
French text

ABSTRACT

Due to the fast growth of social networks, humans have transformed from being general users to creators of network information's by providing reviews, evaluations, and thoughts on social websites expressing their feelings on various topics. Recently, feedback analysis has become important not only for business owners to improve their products based on user feedback, but also for users to help them select the most suitable products by benefiting from other's experiences. Extracting and identifying human emotional states such as happiness, anger, and worry in texts are targets of emotion analysis due to their importance in providing suggestions for companies and users according to their needs. Although, there has been a lot of work on emotion detection in English text, there is currently lack of research on French text that is because of not existing of French emotion dataset. This paper presents an emotion detection model that integrates the Camembert tokenizer with bidirectional long short-term memory (Bi-LSTM) for emotion detection in French text. The proposed model is trained and validated using a dataset that has been annotated for emotions in French. The proposed model achieved accuracy and an F1-score of 98.66% and 98.66%, respectively, outperforming previous work by 26.36%.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Aya Adel
Department of Computer Science, Faculty of Computers and Artificial Intelligence, Fayoum University
Fayoum, Egypt
Email: aa2580@fayoum.edu.eg

1. INTRODUCTION

Emotion detection, an extension of sentiment analysis, involves discerning emotions such as depression, anger, anxiety, happiness, and sadness, as well as utilizing this insight to help future decision-making [1]. Emotions are integral to human life, affecting decision-making and enriching communication with the world. Identifying an individual's various feelings or emotions, such as joy, sadness, or happiness is referred to as emotion detection or emotion recognition. Emotion detection and analysis are among the most important challenging new issues in the field of natural language processing (NLP). So that detecting a human's emotional status from text data is an active area of research, alongside the identification of emotions from facial expressions and audio recordings. The study of emotions has several applications in different fields, like human-computer interaction, data mining, neuroscience, psychology, e-learning, information filtering systems and cognitive science. The high availability of text sources, such as blogs, customer reviews, social media, and news articles, presents valuable resources to research various aspects in text mining, including emotions [2]. In recent years, automated emotion detection has become a vital area of research [3].

Emotions detection can be detected through physical activities, such as heart rate, hand shaking, perspiring, and voice tones, which also express the emotional state of humans [4]. However, detecting emotions from text is more challenging. Text emotion detection is more complicated due to the introduction of new slang and various ambiguities that arise with each passing day [5]. Emotion analysis approaches based on text are classified into three categories: i) lexicon based, ii) machine and deep learning (DL) based, and iii) hybrid approaches [6]. In lexicon-based approaches, text is classified using a list of manually labeled words. There are two available lexicons namely EmoSenticNet and NRC-Emo Lex. EmoSenticNet [7] is an online lexicon for emotion that includes 13,171 words labeled with emotions such as sadness, joy, disgust, surprise, anger, and fear. While, National Research Council (NRC) emotion word organized lexicon (EmoLex) [8] contains above of 14,000 words labeled into eight categories of emotion. The NRC dataset is translated into more than 40 languages.

Furthermore, machine learning-based (ML) approaches combine ML techniques with textual features to classify the text. Textual features are extracted to provide numerical or categorical representations used as input for ML models [9]. Finally, hybrid approaches combine lexicon and ML techniques to enhance the performance of emotion classification. The size, quality, and language of the data play significant role in the effectiveness of these approaches.

The most widely used ML techniques are support vector machines (SVM), Naive Bayes (NB), and random forests, while DL models like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have also been utilized. Felbo *et al.* [10] introduced a DL approach called DeepMoji, which uses RNNs and attention mechanisms to detect emotion from text. The proposed approach was trained and evaluated using a large dataset of 1246 million English tweets containing about 64 common emojis. The proposed approach achieved great performance on multitasks of emotion classification, such as Amazon Mechanical Turk's sentiment prediction of tweets.

Buechel and Hahn [11] constructed a multiclass balanced emotion dataset called EmoBank consisting of 10,000 English sentences. Then, authors analyzed dimensional emotion performance using different models, including lexicon-based and ML methods. The experimental results proved that the automatic mapping between categorical and dimensional formats using ML techniques is similar to human performance compared to other methods.

Haryadi and Kusuma [12] developed a multi-class emotion detection model based on long short term memory (LSTM), Nested LSTM, and SVM approaches. The proposed model was trained and evaluated using 980,549 and 144,160 tweets for the training and testing datasets, respectively. The depicted experimental results showed that Nested LSTM outperformed other methods by achieving an accuracy of 99.167%.

Machová *et al.* [13] compared different ML, and lexicon-based techniques, such as SVM, NB, and deep neural networks (DNN) for the task of emotion detection. The proposed models were trained using 20,000 posts gathered from the social network’s conversations, containing six emotions: joy, love, sadness, surprise, anger, and fear. The experiments showed that DL model utilizing 1D convolutional layer (Conv1D) and LSTM network achieved the greatest accuracy of about 91%.

Blandin *et al.* [14] analyzed the views and emotions contained in emails to gauge their effect on open and click rates, investigating how analyzing emotions in French emails can boost market success. This study utilized K-means clustering approach to construct the dataset by categorizing the emails based on their emotional content. Emotional embeddings were then employed to explore the relations between email performance metrics, such as click and open rates. Hints on how to write effective email campaigns were then provided. The proposed AdaBoost-based emotion model achieved an accuracy of 72.23%.

Genest *et al.* [15] presented an emotion detection model based on bidirectional encoder representations from transformers (BERT) to detect emotions from French text. Around 14,997 English scripts gathered from the TV show Friends were translated into French for model training. Filtering the dataset and extracting emotional lines resulted in 9,965 samples. For translation, Needleman-Wunsch algorithm was utilized for aligning both English and French scripts. The BERT-based model achieved an F1 score of 67%. Yohanes *et al.* [16] investigated the performance of various RNN DL models, such as LSTM, Bi-LSTM and gated recurrent unit (GRU) for detecting emotions in textual data. On evaluating the proposed models using the ISEAR dataset, GRU-based emotion detection model outperformed the others by achieving an accuracy and F1-score of 60.26% and 59.87%, respectively.

Sharma *et al*. [17] developed a LSTM-based emotion detection system in textual data. This study leveraged the capability of LSTM network to capture long-term dependencies in textual data to make it suitable for emotion classification. A dataset of about 16,000 rows of data, classified into six emotions, was collected and used for training and evaluating the proposed system. The Bi-LSTM model achieved the best performance with an accuracy of 88%.

Essebbar *et al.* [18] utilized pre-trained models like multilingual BERT (mbret), Camembert and Flaubert, and some DL techniques to detect aspect based sentiment analysis (ABSA). On evaluating the proposed models using SemEval2016, experiments showed that Flaubert French pre-trained model outperformed other models by achieving an accuracy and F1-score of 84.83% and 73.84%, respectively. Adoma *et al.* [19] proposed a two-stage architecture utilizing BERT and Bi-LSTM as encoder and decoder for emotion detection. On evaluating the proposed architecture using ISEAR dataset, a F1-score of 73% was achieved.

Abas *et al.* [20] proposed a hybrid emotion detection model based on BERT and CNN. BERT was utilized to extract the word context. Then, BERT output was passed to CNN for emotion prediction. On evaluating the proposed model using ISEAR, an accuracy and F1-score of 75.8% and 76% were achieved, respectively. When using semeval 2019 task3 dataset, the proposed model achieved an accuracy of 94.7%, and an F1-score of 94%.

Bharti *et al.* [21] proposed a hybrid DL and ML text-based emotion recognition model. The proposed model utilized three different datasets, namely, ISEAR, WASSA, and the emotion-stimulus dataset. The hybrid proposed model utilized word2vec as embedding layer. The resulted embedding vector was fed into both CNN and bidirectional gated recurrent unit (Bi-GRU), which acted as encoders after removing their last layer. Finally, SVM was used as a classifier. The proposed hybrid model achieved a precision, a recall, an F1 score, and an accuracy of 82.39%, 80.40%, 81.27%, and 80.11%, respectively.

Alvi *et al.* [22] presented a novel hybrid method based on ML techniques and recursive data munging module to extract user's opinions from a Twitter dataset regarding warming and weather changes. The proposed model achieved an accuracy of 86.18%. Deshpande and Paswan [23] proposed a hybrid method for emotion detection from Twitter posts, which is based on a rule-based component and ML algorithms. The experimental results showed that the proposed method achieved better performance compared to the older Unison model with an average accuracy of 88.39%.

In this paper, a novel model for emotion analysis, integrating the Camembert tokenizer with Bi-LSTM classifier, is proposed to detect emotions from French text. The main contributions of this paper are summarized as follows:

- Constructing an annotated French emotion detection dataset from an existed sentiment dataset.
- Developing a novel emotion detection model to detect emotions from French text, which overcomes the limitations of other models.
- Outperforming other French emotion detection models by achieving an accuracy of 98.66%, with an increase of 26.36%.

The remaining of this paper is organized as follows phases of the proposed emotion detection model are described in section 2. The obtained experimental results are presented and discussed in section 3, and section 4 introduces the conclusions and future work.

2. METHOD

This section presents the proposed model phases, which leverages the power of Camembert and Bi-LSTM. Camembert is based on the BERT technique, which is specifically designed for French text. It was trained using a large French corpus and achieved state-of-the-art performance in various tasks, such as questions answering, sentiment analysis, and part-of-speech tagging in French text [24]. Bi-LSTM is one of RNN architectures, which enhances the traditional LSTM network by integrating bidirectional processing. This approach replaced each word in sentence with its corresponding integer value from the vocabulary index. Consequently, an input sentence can be converted to a vector with size equal to the number of words it contains [3]. The proposed model uses the tokenizer of Camembert alongside Bi-LSTM for emotions detection in French text. It consists of four phases: data collecting and cleaning, preprocessing, emotion identification, and testing phases as shown in Figure 1.

2.1. Data collection and cleaning phase

Since there is no available French emotion dataset. The dataset utilized in this paper was constructed using the following steps:

a. Downloading French sentiment analysis dataset from Kaggle [25].
b. Annotating about 50,000 records from the downloaded dataset as emotion dataset consisting of 5 emotion labels (happy, stratification, angry, sad, and worry), manually.
c. Removing duplicated rows and empty rows.

Figure 2 shows emotion distribution in the constructed dataset, which is 25.9%, 14%, 10.2%, 23%, and 27% records for happiness, sadness, angry, satisfaction, and worry classes, respectively.

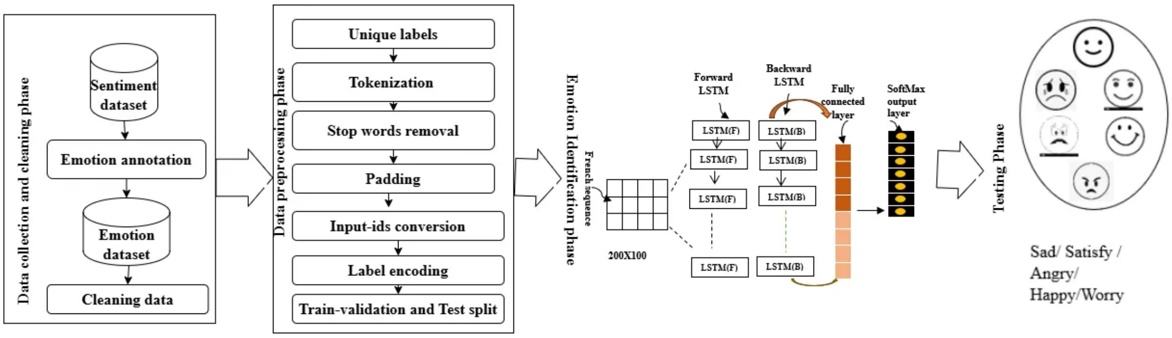

Figure 1. Proposed emotion detection model

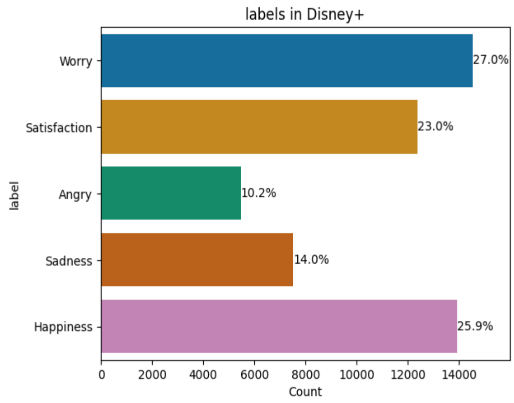

Figure 2. Emotions distribution

2.2. Data preprocessing phase

Preprocessing is a crucial phase for text data to eliminate noise in order to help classifiers achieve higher performance and effectiveness. Steps in the preprocessing phase can be different based on the problem and the used data. In this paper, the following steps are utilized and performed on text before training or evaluating the model:

a. Unique labels: the data is manually annotated, which may result in misspelling in labels, a python dictionary is used to correct these errors.
b. Tokenization: this step involves dividing text into smaller units called tokens, without understanding the meaning or relationships of words. These tokens are then used as input for the next steps. In this work, the Camembert tokenizer [23] is utilized. For example, given the following input: text="J'aime le chocolat.", MAX_LENGTH=10. The Camembert tokenizer divides the text into the following tokens: tokens = ['▁J', "'", 'aime', '▁le', '▁chocolat', '.']. This is because of Camembert tokenizer is aware of the linguistic features of the French language. It provides superior tokenization for French text compared to the Keras Tokenizer.
c. Stop words removal: this step is responsible for removing stop words from the tokenized text. Stop words are the general words that often do not contribute much to the general meaning of a sentence. In this work, natural language toolkit’s (NLTK) list of stop words is used.
d. Padding: the maximum sequence length (MAX_LENGTH) is set to 200. As shown in Figure 3, some sentences have a length longer than 200. To uniform the input, this step padded all tokens to a length of 200.
e. Input-IDs conversion: in NLP tasks, this step converts textual data into input IDs, a form that the models such as NN can process. Each token is replaced with a specific number. For example, tokenizing a sentence and replacing each token with its corresponding ID create a sequence of integers the represents the original sentence.
f. Label encoding: this step converts the class/category labels into numeric values for ML processing. The labels-sad, angry, worry, happy, and satisfaction, are encoded as 0, 1, 2, 3, and 4 respectively.
g. Train-validation and test split: in this step, dataset is divided into training and testing datasets to train and evaluate model performance.

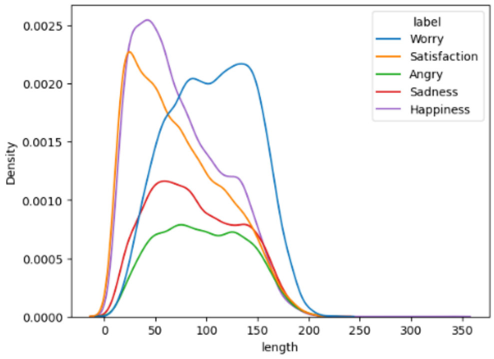

Figure 3. Emotions sentences length

2.3. Emotion identification phase

The proposed classification model is designed to detect emotion from French text, where it accepts a train data to train on to predict the class or emotion to which the text has. The proposed model has four main components:

a. Embedding layer: the words or tokens are classically represented as one-hot vectors, where every word is assigned a unique index in a vocabulary. However, one-hot vectors are high-dimensional and sparse, making them less efficient for most ML algorithms. The embedding layer aims is to address this issue by learning a dense representation for each word or token. During the training process, the embedding layer adjusts its parameters to map each word to a continuous-valued vector of lower dimensionality. The resulting embeddings effectively capture the semantic relationships between words, such as similarity and context. These learned representations can be used as input to downstream tasks such as sentiment analysis, machine translation, text classification, and more.
b. Bidirectional LSTM layer: this layer captures sequential information in the input embeddings. It receives the input text in shape of (batch size, MAX_LENGTH, embedding size), and produces as output with shape of (batch size, hidden_size1).
c. Dense layers: these layers perform non-linear transformations on the LSTM output to learn complex representations. They receive the input text in shape of (batch size, hidden_size1), and produces output with shape of (batch size, dense size).
d. Output layer: this layer computes the final output predictions for every class using the SoftMax activation function. It receives the input text in the shape of (batch size, dense size), and produces output with shape of (batch size, Num classes).

2.3.1. Training model

After conducting several experiments to select the parameters, which achieve the best performance, the proposed model is trained and validated using training and validation datasets. The number of epochs, batch size, learning rate for Adam optimizer are set to 400, 64, and 0.001, respectively.

2.4. Testing phase

Finally, the proposed model is tested and evaluated using testing dataset. The experimental results will be discussed in section 3. The utilized performance metrics for model evaluation are accuracy, and F-measure, recall and precision, which are commonly used in classification problems.

3. RESULTS AND DISCUSSION

This section reports and discusses the experimental results in details to show the superiority of the proposed model compared to the state-of-art models using various evaluation metrics. Experiments were performed on Kaggle platform and developed using Python.

3.1. Dataset

As mentioned above, the used dataset was constructed utilizing a French sentiment dataset. Then, about 57000 records were annotated manually in five categories, namely, (sad, angry, worry, happy, and stratification). Samples of data are shown in Figure 4. Column 1 presents French statement and column 2 presents emotion label. The dataset is divided into 60%, 20%, and 20% for training, validation, and testing datasets respectively.

text	label
Quel est le problème avec mes blips!?! Ils ne ...	Worry
quelle? non! C'est le même temps que david arc...	Worry
J'ai passé une bonne nuit de repos ...	Satisfaction
Mais ils ont frappé doug pour réveiller dans s...	Worry
Je me suis énervé	Angry

Figure 4. Dataset samples

3.2. Experiments results

After investigated several experiments, the depicted results shows that integration between Camembert tokenizer and Bi-LSTM classifier achieved the best performance because the Camembert tokenizer is aware of French linguistic patterns. As shown in Table 1, the proposed model utilizing Camembert tokenizer achieved an accuracy of 98.66%, an F1-measure of 98.66%, a recall of 98.66%, and a precision of 98.66%. On the other hand, using the same model utilizing Keras tokenizer achieved an accuracy of 81.4%, an F1-measure of 81.3%, a recall of 81.52%, and a precision of 81.25%.

The accuracy and loss for both training and validation of the model using Keras tokenizer are shown in Figures 5 and 6. As depicted, the model shows no improvement after a limited number of epochs. Figures 7 and 8 show the accuracy and loss for both training and validation of the model using Camembert tokenizer. As illustrated, the model shows a significant improvement over the epochs. The confusion matrix is shown in Figure 9, illustrating that model is able to accurately predict emotions by achieving an accuracy of 97%, 98%, 99%, 99%, and 99% for sadness, angry, satstification, and happiness emotions, respectively. However, the sad emotion was confused by other emotions.

Table 1. Performance of the proposed model using Keras tokenizer and Camembert tokenizer

Evolution matrix	Using Camembert tokenizer	Using Keras tokenizer
Accuracy	98.66	81.4
F-score	98.66	81.3
Precision	98.66	81.52
Recall	98.66	81.25

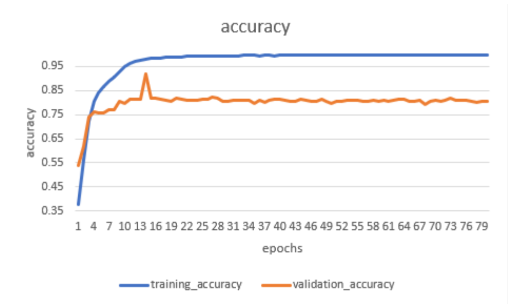

Figure 5. Training and validation accuracies using Keras tokenizer

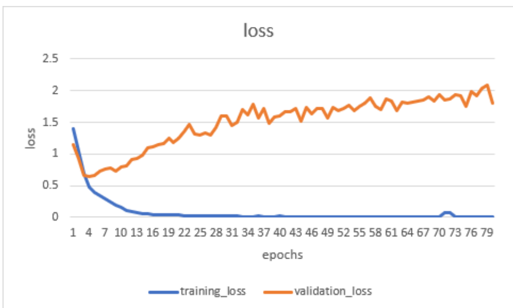

Figure 6. Training and validation losses using Keras tokenizer

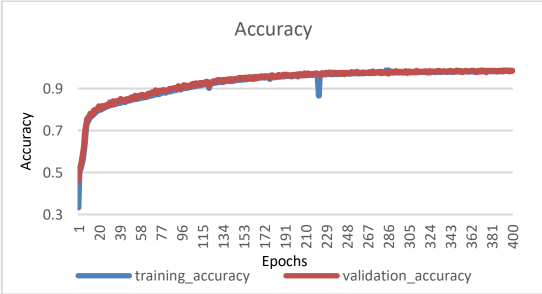

Figure 7. Accuracy and validation accuracy using Camembert tokenizer

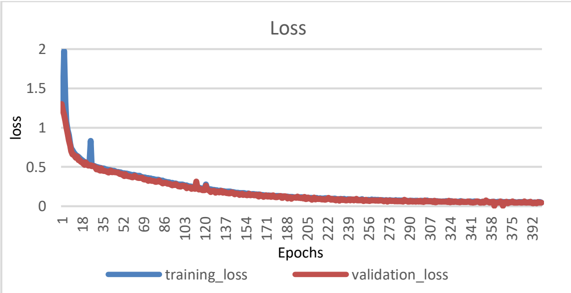

Figure 8. Accuracy and validation loss using Camembert tokenizer

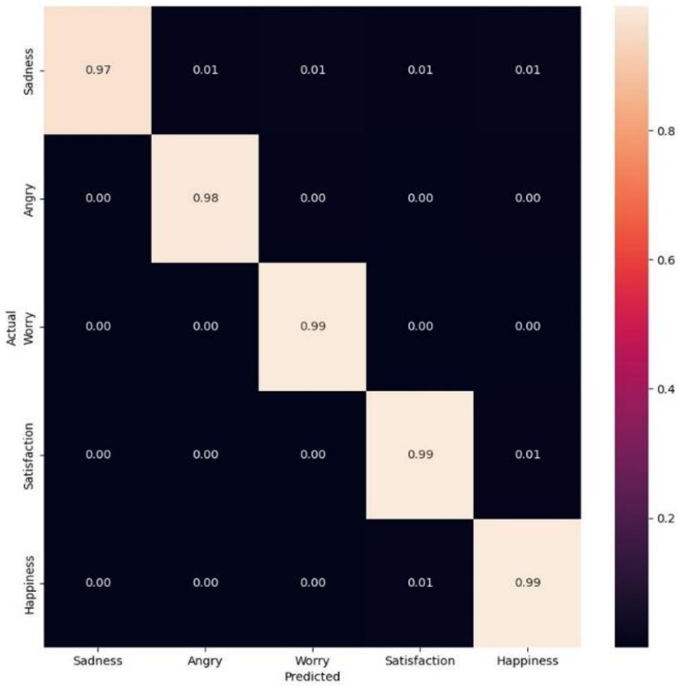

Figure 9. Confusion matrix using Camembert tokenizer

The innovation of the proposed French model is its training and testing using French text. Previous work on emotion detection in French text hasn't reached the proposed model accuracy as shown in Table 2. For example, in Genest *et al.* [15] built their dataset using scripts translated from English to French and achieved an F1-score of 67%. Meanwhile, in Blandin *et al.* [14] classified the content of e-mails and achieved an F1-score of 72.3%. The proposed model outperforms both, with an improvement of 26.36%.

Table 2. State of art French text

Model	F1-score (%)	Accuracy (%)	Recall (%)	Precision (%)	Dataset	Method	Number of classes
The proposed model	98.66	98.66	98.66	98.66	Annotated dataset	Proposed model	5
[14]	72.3	-	72.3	72.3	French mail contents	AdaBoost	2
[15]	67	-	-	-	Translated scripts	Classifier depends on BERT	4

4. CONCLUSION

This paper presented a novel model for emotion detection in French text that integrates the power of Camembert tokenizer, an embedding layer and a BI-LSTM classifier. As mentioned, there is no available French emotion dataset. Therefore, in this study, an emotional French dataset was manually generated by transforming a sentiment dataset into emotion one. The experimental results showed that Camembert tokenizer improved the model performances as it built especially for French text. The proposed emotion detection model achieved an accuracy of 98.66%, outperforming previous work by 26.36%. For future work, several challenges and research directions could be considered, such as studying the effect of considering various aspects of the text on emotion detection accuracy and conducting more experiments with datasets in low-resource languages, such as Arabic, where there has been less research and performance analysis.

ACKNOWLEDGEMENTS

We would like to thank MEAPAL data Science @ CORELIA https://www.corelia.ai/ for their great collaboration and support.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Aya Adel	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
Shereen A. Taie	✓					✓	✓	✓		✓	✓	✓	✓	
Esraa Elhariri	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	
Mohamed Hasan Ibrahim	✓					✓				✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

Derived data supporting the findings of this study are available from the corresponding author [A.A] on request.

REFERENCES

[1] S. Kusal, S. Patil, K. Kotecha, R. Aluvalu, and V. Varadarajan, “AI Based Emotion Detection for Textual Big Data: Techniques and Contribution,” *Big Data and Cognitive Computing*, vol. 5, no. 3, p. 43, Sep. 2021, doi: 10.3390/bdcc5030043.

[2] A. R. Murthy and K. M. A. Kumar, “A Review of Different Approaches for Detecting Emotion from Text,” *IOP Conference Series: Materials Science and Engineering*, vol. 1110, no. 1, p. 012009, Mar. 2021, doi: 10.1088/1757-899X/1110/1/012009.

[3] P. Nandwani and R. Verma, “A review on sentiment analysis and emotion detection from text,” *Social Network Analysis and Mining*, vol. 11, no. 1, p. 81, Dec. 2021, doi: 10.1007/s13278-021-00776-6.

[4] B. Kratzwald, S. Ilić, M. Kraus, S. Feuerriegel, and H. Prendinger, “Deep learning for affective computing: Text-based emotion recognition in decision support,” *Decision Support Systems*, vol. 115, pp. 24–35, Nov. 2018, doi: 10.1016/j.dss.2018.09.002.

[5] A. A. Maruf, F. Khanam, M. M. Haque, Z. M. Jiyad, M. F. Mridha and Z. Aung, “Challenges and Opportunities of Text-Based Emotion Detection: A Survey,” in *IEEE Access*, vol. 12, pp. 18416-18450, 2024, doi: 10.1109/ACCESS.2024.3356357.

[6] F. A. Pozzi, E. Fersini, E. Messina, and B. Liu, *Sentiment Analysis in Social Networks*. 2016.

[7] S. Poria, A. Gelbukh, A. Hussain, N. Howard, D. Das, and S. Bandyopadhyay, “Enhanced SenticNet with Affective Labels for Concept-Based Opinion Mining,” *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 31–38, 2013, doi: 10.1109/mis.2013.4.

[8] S. M. Mohammad and P. D. Turney, “Crowdsourcing a Word–Emotion Association Lexicon,” *Computational Intelligence*, vol. 29, no. 3, pp. 436–465, Aug. 2013, doi: 10.1111/j.1467-8640.2012.00460.x.

[9] M. Mohan, A. K. Abhinav, A. Ashok, A. V. Akhil, and P. R. Achinth, “Depression Detection using Facial Expression and Sentiment Analysis,” *2021 Asian Conference on Innovation in Technology, ASIANCON 2021*, 2021, doi: 10.1109/ASIANCON51346.2021.9544819.

[10] B. Felbo, A. Mislove, A. Søgaard, I. Rahwan, and S. Lehmann, “Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm,” *EMNLP 2017-Conference on Empirical Methods in Natural Language Processing, Proceedings*, pp. 1615–1625, 2017, doi: 10.18653/v1/d17-1169.

[11] S. Buechel and U. Hahn, “EMOBANK: Studying the impact of annotation perspective and representation format on dimensional emotion analysis,” *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017-Proceedings of Conference*, vol. 2, pp. 578–585, 2017, doi: 10.18653/v1/e17-2092.

[12] D. Haryadi and G. P. Kusuma, “Emotion detection in text using nested Long Short-Term Memory,” *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, pp. 351–357, 2019, doi: 10.14569/ijacsa.2019.0100645.

[13] K. Machová, M. Szabóova, J. Paralič, and J. Mičko, “Detection of emotion by text analysis using machine learning,” *Frontiers in Psychology*, vol. 14, Sep. 2023, doi: 10.3389/fpsyg.2023.1190326.

[14] A. Blandin, F. Saïd, J. Villaneau, and P.-F. Marteau, “Automatic Emotions Analysis for French Email Campaigns Optimization,” in *Centric 2021*, Oct. 2021.

[15] P. Y. Genest, L. W. Goix, Y. Khalafaoui, E. Egyed-Zsigmond, and N. Grozavu, “French translation of a dialogue dataset and text-based emotion detection,” *Data and Knowledge Engineering*, vol. 142, Nov. 2022, doi: 10.1016/j.datak.2022.102099.

[16] D. Yohanes, J. S. Putra, K. Filbert, K. M. Suryaningrum, and H. A. Saputri, “Emotion Detection in Textual Data using Deep Learning,” in *Procedia Computer Science*, 2023, vol. 227, pp. 464–473, doi: 10.1016/j.procs.2023.10.547.

[17] J. Sharma, A. Pandey, and P. Nanda, “Emotion Detection System Using Long Short Term Memory (LSTM),” *International Journal of Scientific Research and Engineering Development*, vol. 7, no. 3, pp. 737–743, 2024.

[18] A. Essebbar, B. Kane, O. Guinaudeau, V. Chiesa, I. Quénel, and S. Chau, “Aspect Based Sentiment Analysis using French Pre-Trained Models,” in *Proceedings of the 13th International Conference on Agents and Artificial Intelligence*, vol. 1, 2021, pp. 519–525, doi: 10.5220/0010382705190525.

[19] A. F. Adoma, N.-M. Henry, W. Chen, and N. R. Andre, “Recognizing Emotions from Texts using a Bert-Based Approach,” in *2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, Dec. 2020, pp. 62–66, doi: 10.1109/ICCWAMTIP51612.2020.9317523.

[20] A. R. Abas, I. Elhenawy, M. Zidan, and M. Othman, “Bert-cnn: A deep learning model for detecting emotions from text,” *Computers, Materials and Continua*, vol. 71, no. 2, pp. 2943–2961, 2022, doi: 10.32604/cmc.2022.021671.

[21] S. K. Bharti *et al.*, “Text-Based Emotion Recognition Using Deep Learning Approach,” *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–8, Aug. 2022, doi: 10.1155/2022/2645381.

[22] M. B. Alvi, N. A. Mahoto, M. Alvi, M. A. Unar, and M. A. Shaikh, “Hybrid Classification Model for Twitter Data - A Recursive Preprocessing Approach,” in *2018 5th International Multi-Topic ICT Conference (IMTIC)*, Apr. 2018, pp. 1–6, doi: 10.1109/IMTIC.2018.8467221.

[23] A. Deshpande and R. Paswan, “Real-Time Emotion Recognition of Twitter Posts Using a Hybrid Approach,” *Ictact Journal on Soft Computing*, vol 10, no. 10, p. 4, Jul. 2020, doi: 10.21917/ijsc.2020.0302.

[24] L. Martin *et al.*, “CamemBERT: A tasty French language model,” in *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Jul. 2020, pp. 7203–7219, doi: 10.18653/v1/2020.acl-main.645.

[25] Hbaflast, “French Twitter Sentiment Analysis.” Kaggle, [Online]. Available: https://www.kaggle.com/datasets/hbaflast/french-twitter-sentiment-analysis/data. (Accessed: Aug. 12, 2023).

BIOGRAPHIES OF AUTHORS

Aya Adel received the B.Sc. degree in computer science from Faculty of Computers and Artificial Intelligence, Fayoum Egypt, in 2017 and is currently enrolled master point in computer science at the same university. She is a Teaching Assistant at the Faculty of Artificial Intelligence, Egyptian Russian University, Cairo, Egypt. She has worked as mobile developer remotely for green-hat company, worked as teaching assistant at must university Cairo, Egypt from expertise includes programming languages like Python, C, C++, C#, and Java, along with proficiency in libraries and frameworks such as NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn, TensorFlow, and Keras. She has experience in mobile development using native platforms and cross plat form like flutter. Certificated as Artificial Intelligence Practitioner instructor by IBM. She has 3 publications in IEEE conferences. She can be contacted at email: aa2580@fayoum.edu.eg or aya-adelmahmoud@eru.edu.eg.

Shereen A. Taie is an Artificial Intelligence Professor in Computer Science Department, Faculty of Computers and Information, Fayoum University, Egypt, and Vice Dean for Postgraduate Studies and Research, Faculty of Computers and Artificial Intelligence, Fayoum University, Egypt. Head of Center of Electronic Courses Production, Fayoum University, Egypt. She held an M.S. in 2006 in Computer Science from Computer and Mathematics Department Faculty of Science, Cairo University, Egypt, and a Ph.D. in 2012 in Computer Science from Computer and Mathematics Department Faculty of Science, Cairo University, Egypt. She was awarded a B.Sc. degree in Computer Science by Faculty of Science, Cairo University in 1996. She is the Vice Dean for Education and Students Affairs, Faculty of Computers and Information, Fayoum University, Egypt. She is the head of "Artificial Intelligence Research Group" in Faculty of Computer and Information, Fayoum University, with overall 20 years of in-depth experience in teaching field along with researching experience gained during participation in +39 publications that had been issued in highly ranked international journals. Her H-index is 14.03, her teaching interest and research interest in artificial intelligence, data analysis, data science, image processing, machine learning , algorithms, and cloud computing. She is a certified auditor in the National Authority for Quality Assurance and Accreditation, and a reviewer for research projects in the National Telecom Regulatory Authority (NTRA), Egypt. She can be contacted at email: sat00@fayoum.edu.eg or shereentaie@yahoo.com.

Esraa Elhariri received the Ph.D. degree in computer science from the Faculty of Computers and Artificial Intelligence, Fayoum University, Egypt, in 2020, and the M.Sc. degree from FCI-Cairo University, Egypt, in 2015. Her Ph.D. topic was about IoT-based techniques for historical and cultural heritage preservation. She is currently a Lecturer with the Faculty of Computers and Artificial Intelligence, Fayoum University. She has published several articles in international journals and peer-reviewed international conference proceedings along with chapters. Her research interests include machine learning, image and signal processing, bio-inspired optimization, wireless sensors network cloud computing security, and IoT. She can be contacted at email: emh00@fayoum.edu.eg.

Dr. Mohamed Hasan Ibrahim is an accomplished data scientist and business analyst with 30 years of IT consulting experience in the Middle East. He holds a Ph.D. in data science and is currently an Associate Professor in the field. With expertise in business intelligence, process mining, and digital transformation, he has made significant contributions to complex data science projects across various industries. He has published over 30 papers and supervised multiple theses, demonstrating his commitment to academic excellence. He is a data science Lab Manager @MEAPAL/CORELIA.Ai, A RICOH company. He is diverse expertise and leadership have earned him recognition in the industry. He can be contacted at email: Mhi11@fayoum.edu.eg or MIB@CORELIA.AI.