Detecting Emotions Through Text Using Long Short-Term Memory (LSTM)

William Viriya Halim
School of Computer Science
BINUS University
Jakarta Barat, Indonesia
william.halim005@binus.ac.id

Christhoforus Delon Yudistira
School of Computer Science
BINUS University
Jakarta Barat, Indonesia
christhoforus.yudistira@binus.ac.id

Jurike V. Moniaga
School of Computer Science
BINUS University
Jakarta Barat, Indonesia
jurike@binus.edu

Ghinaa Zain Nabiilah
School of Computer Science
BINUS University
Jakarta Barat, Indonesia
ghinaa.nabiilah@binus.ac.id

***Abstract*—As time passed, humans tend to speak through texts rather than talking face to face due to its convenience. Although this is not a bad thing, this also causes some problems. An example of the problem is misunderstanding. The recipients can't see the texter's expression, therefore they might not understand fully the context behind the texts. In this study, we aim to fix that problem by building a text to emotion classification model with the available resources. The model we used is a deep learning LSTM model with a total of 7 classifying emotions consists of joy, fear, anger, sadness, disgust, shame, guilt. The model used provided us an accuracy of 71.22% in classifying the different emotions. The accuracy of this model is classified good because the majority of the text that goes through the model can be predicted well. For our future research, we hope to improve the model we use by combining existing model with other models to increase the training value of our model so that we can reduce the loss value and increase the accuracy value of our model with hope it reaching above 80-90%.**

Keywords— LSTM, LSTM Model, AI, Computer Science, Emotion Detection, Text Detection, BINUS, Deep Learning, Classification

I. INTRODUCTION

By definition, emotional validation is the ability to acknowledge the emotions felt. Validating feelings does not mean confirming that you are right, but admitting that you are feeling the emotion. Emotional validation is needed by most people who experience difficulty in controlling their emotions, because in some cases emotions can turn into negative definitions. Being able to control emotions is an advantage that humans possess, because by being able to control it we can understand ourselves and the environment around us much better. The inability to control one's emotion often causes people to have emotional strain.

Mental illness or emotional strain has been a problem for a long time, and most of the time, the people having these problems don't really express themselves very well People who are experiencing mental illness or emotional distress frequently resort to online forums for assistance or a sense of human intimacy and understanding instead of seeking professional help[1]. This illustration demonstrates how challenging it may be to identify and categorize various emotions in the context of sentiment analysis[1].

Nowadays, with the ever increasing widespread use of the internet, people tend to prefer chatting through texts rather than face-to-face interactions which makes it difficult for the recipient to understand the texter's emotions since they can't read or see the texter's expression. This often causes a lot of misunderstanding. For this reason, difficulties arise in determining the required response due to our lack of understanding of the emotions displayed through text. Therefore, the use of deep learning with LSTM models is very necessary and can be used to find a solution, by applying training from various datasets so that this LSTM model can work to determine and display emotions from text with fairly high accuracy.

Because Word2Vec can identify comparable words, LSTM with Word2Vec word embedding is frequently used[2]. Due to its ability to extract semantic links between words and co-occurrence, Glo-Ve word embedding is also frequently utilized[2]. The purpose of this work is to detect emotions with seven different emotional labels (joy, fear, rage, sorrow, disgust, humiliation, and guilt) using LSTM with W2V word vector pretrained on Wikipedia. It is anticipated that this research will result in an application or program that can use texts to categorize emotions into groups such as joy, fear, rage, sadness, contempt, shame, and guilt. Making decisions in a variety of industries, including business, education, and hiring new staff, can benefit from the findings of emotional detection[2].

II. LITERATURE REVIEW

An example of an LSTM embedded model for depression analysis has been given. Singh, J. [3]. presented a method for the investigation and identification of depression based on embedded long short-term memory (LSTM). It made use of natural language processing methods to offer a revolutionary framework that predicted the user's sentimental and emotional state depending on how the user interacted with the created model. Along with the practical application of the suggested strategy, significant outcomes were also calculated, including accuracy, precision, recall, and F1 score.

Reviews of TBED that showcase S. Kusal's techniques, strategies, features, etc. [4]. The need for digital web media has grown recently. Using digital web platforms, this article offers a thorough analysis of the literature on tuberculosis-associated. Artificial intelligence methods have been used extensively in digital internet media analysis. The results of a systematic review of the literature on TBED techniques are presented in this study.

S. K. Bharti et al. [5] used a hybrid approach that included machine learning and deep learning techniques to recognize emotions in text. This work presents a text-based paradigm for emotion recognition. The suggested model combines machine learning and deep learning techniques. Three datasets—WASSA, ISEAR, and the EmotionStimulus dataset—are combined in this suggested hybrid technique. The suggested approach has numerous benefits, including the ability to operate on multiple text sentences, tweets, dialogs, keywords, and readily identifiable vocabulary words of emotions.

D. Kher and K. Passi, "Twitter as dataset in emotion classification." [6]. In this study, Twitter data was analyzed to categorize emotions. This challenge has been characterized as multi-label emotion classification because every tweet is linked to numerous emotions rather than just one. The multi-label emotion classification problem was solved using well-known machine learning classifiers and a GRU-based recurrent neural network with Adam and RmsProp optimizer.

A thorough summary of the numerous obstacles in the field of TBED is intended to be very beneficial to upcoming scholars who are interested in this area. Al Maruf, A. [7]. Stages of text preparation are required to improve the accuracy of text-based emotion recognition. This survey described the text's preprocessing processes, including crucial procedures like text cleaning and normalization. This study has looked at lexicons and datasets that are accessible to the public. There is currently no ideal method for text-based emotion recognition, despite several studies being done in a variety of domains and applications.

D. Haryadi and G. Putra Kusuma conducted a study on emotion detection from text, as mentioned in their publication [8]. After conducting a thorough discussion and evaluation, it was determined that LSTM, Nested LSTM, and SVM methods are suitable for multi-class emotion detection. Among these methods, Nested LSTM achieved the highest accuracy of 99.167%. Although LSTM had a slightly lower accuracy of 99.154%, it demonstrated better average performances in terms of precision, recall, and f1-score, with values of 99.22%, 98.86%, and 99.04% respectively.

N. Aslam [9] analyses tweets about cryptocurrencies using sentiment analysis and emotion detection. Cryptocurrency sentiment analysis is potentially important since it is frequently used to forecast market prices, which requires very accurate sentiment classification. Tweets are taken from TwitterTM for the experiments, and TextBlob and Text2Emotion are used to annotate the dataset for sentiments and emotions, respectively. Furthermore, the machine learning models employ BoW, TFIDF, and Word2Vec features as feature extraction methods. The findings show that machine learning models outperform TF-IDF models when using BoW features. And Word2Vec.

LSTM in both directions for text classification. A bi-directional LSTM language model based on text emotion recognition is proposed by Z. Wang [10] for the English short text classification challenge. In order to estimate the final emotional output, this article employs Word Embedding to encode words, express text as word feature representation and character feature representation, and mix content words with emotion functional words. According to the experimental data, this model has a high multi-classification detection accuracy of up to 64.09% and a quick rate of convergence.

Batbaatar, E. [11]. This study investigated the emotion recognition from text using a novel model named SENN, which is applied to ten real-world datasets. SENN is a hybrid network that comprises of a CNN-based emotion encoder and a BiLSTM-based semantic encoder. CNN is made to efficiently extract emotional information, while BiLSTM is made to capture contextual information for the SENN model.

Sentiment analysis of text with fewer layers in neural networks. Chaudhari, N. [12]. This study offers a system that analyses written sentiments and classifies face expressions using technologies like artificial intelligence and machine learning, which are based on the principles of deep learning. With the aid of four widely used datasets, the proposed system enables neural networks with fewer layers to compete with or rather, outperform much more complex and deeper networks in FER. This method of emotion sensing can be thought of as one of the most widely used AI and pattern analysis applications.

M. J. Basha in another work [13]. The field has undergone a revolution thanks to pretrained language models, contextual word embeddings, attention mechanisms, and multimodal learning techniques, which have improved machine comprehension of human language. These advancements have enhanced already-existing applications and created new opportunities in machine translation, question-answering systems, sentiment analysis, and information retrieval. But there are still difficulties. The proper application of NLP technologies, bias mitigation, and ethical issues are important topics that necessitate continuous research and improvement.

A 98% accurate text classification with CNN and LSTM. [14] S. Velampalli. The LSTM NN and Standard fully connected NN models are trained using the embeddings. The text categorization accuracy reported in this article was approximately 98% for both algorithms. Conversely, the accuracy of both models dropped sharply to 70% when emojis that weren't in the training set were used to create the validation set. For improved scalability, the models were also trained via the distributed training technique rather than the conventional single-threaded model. By employing the distributed training methodology, the models managed to decrease the runtime by approximately 15% while maintaining accuracy levels.

Haryadi, D. [15]. presented a study on text-based emotion recognition. LSTM, Nested LSTM, and SVM techniques can be applied for multi-class emotion recognition, according to the analysis and testing done in the preceding section. Out of the three techniques, nested LSTM has the highest accuracy (99.167%). This precision is comparable to that of the LSTM, which has an accuracy of 99.154%, without any difference. With average performances of 99.22%, 98.86%, and 99.04% in precision, recall, and f1-score, respectively, LSTM outperforms other neural networks.

Extracting hidden features with the use of CNN and Bi-LSTM, by J.L. Wu. [16]. suggested using a deep learning model to tag emotions in social messages used in mental health. To extract the hidden features, the proposed

Bidirectional LSTM-CNN integrates convolutional neural networks, word embedding, and long short-term memory networks. The word hidden features are derived from word embeddings, and the major decision hidden features are derived from CNN's prior hidden features. Experimental results show that the proposed deep learning model significantly improved performance over other conventional models.

This paper includes studies on sentiment analysis from various data sources, by S. Ahmad. [17]. We have utilized the Attention-based C-BiLSTM model, a deep learning approach, to classify English poetry texts into multiple emotion classes. Following data acquisition, preprocessing, feature representation, feature extraction, feature encoding, context information production, and classification, this dataset is then run through the following modules. The results show that, when compared to state-of-the-art studies, the suggested approach performed best in terms of better (0.88%) precision, (0.88) recall, (0.88%) f-measure, and (88%) accuracy.

Focuses on the use of the CNN-BiLSTM algorithm for emotion recognition in an intelligent learning environment by X. Lu. [18]. The current techniques of learning have started to shift as a result of the Internet's rapid development. Based on the principle of emotional interaction amongst learners in an intelligent learning environment, the CNN-BiLSTM algorithm is used to recognize learner's emotions in real time, facilitating calm and productive learning. According to the experimental findings, the CNN-BiLSTM algorithm presented here has an accuracy of 98.75%, which is at least 3.15% higher than other comparison algorithms' accuracy, and a recall rate that is at least 7.13% higher than other algorithms.

In this paper refers to the background and foundational concepts of deep learning, by L. Alzubaidi. [19]. It is required to include a brief commentary after compiling all of the pertinent information obtained throughout this in-depth investigation. After that, an itemized analysis is provided to show the future orientation and to wrap up our review. including its inspiration from the human brain and its reliance on large datasets for mapping inputs to specific labels.

This paper includes considered modalities represent promising information sources, by G. Placidi. [20]. give a reliable interaction and the time limits needed to incorporate the identified emotion into the robot's behaviour adaption. Furthermore, a large portion of the used dataset originated from generic HMI research, making them unsuitable for real-world emotion recognition applications. There is still a demand for true HRI datasets. In fact, it's possible that the robot's field of vision is not lined up with the pictures included in the dataset.

III. Methodology

A. *Data Collection and Labelling Data*

The dataset used to this study to train the model was an ISEAR (International Survey on Emotional Antecedents and Reactions). The ISEAR for data training consist of 7.6k lines of texts. Each texts has an emotion labelling consists of 7 different emotions such as joy, fear, anger, sadness, disgust, shame, guilt. An example of the dataset in this study can be seen in TABLE I.

TABLE I.

Emotion	***Text***
Joy	On days when I feel close to my partner and other friends. When I feel at peace with myself and also experience a close contact with people whom I regard greatly.
Fear	Every time I imagine that someone I love or I could contact a serious illness, even death.
Anger	When I had been obviously unjustly treated and had no possibility of elucidating this.
Sadness	When I think about the short time that we live and relate it to the periods of my life when I think that I did not use this short time.
Disgust	At a gathering I found myself involuntarily sitting next to two people who expressed opinions that I considered very low and discriminating.
Shame	When I realized that I was directing the feelings of discontent with myself at my partner and this way was trying to put the blame on him instead of sorting out my own feelings.
Guilt	I feel guilty when I realize that I consider material things more important than caring for my relatives. I feel very self-centered.

B. *Preprocessing*

Preprocessing is the next step following labelling data so that emotions can be identified. Machine learning performs better when data preparation is used, as stated in reference [2]. To prepare the text for classification, preprocessing can be used to remove unnecessary information and noise from the data[2]. Preprocessing will clean up the data, improving the accuracy of word vector and classifier creation[2]. Preprocessing in this study involves label conversion, tokenization, and padding. Although stop word removal is an option, lemmatization and stemming are not necessary because word2vec allows words with the same stem to have distinct meanings.

- Tokenization is the process of breaking down a written document into discrete components called tokens [3]. A token may consist of a single letter, such as punctuation, or a word or word fragment[3]. Through it, we can train the designed model more effectively[3]. Since it is preferable to examine the data independently, we may simply discern the meaning of the text by analyzing its words[3]. An example of tokenization can be seen in TABLE II.

TABLE II.

Tokenization Process	***Text***
Text before tokenization	"When I found a bristle in the liver paste tube."
Text after tokenization	["When", "I", "found", "a", "bristle", "in", "the", "liver", "paste", "tube"]

- Every input has the same length thanks to padding. We set a maximum word count for our texts, and the input size to our model must be fixed. To maintain the same input length, we must pad with zeros.

- Label conversion consist of converting the labels into integers and categorizing them. First, we convert the labels from the datasets from strings into integers, then we can categorize them accordingly. An example of converting labels into integers can be seen in TABLE III.

TABLE III.

Emotion	*Integers*
Joy	0
Fear	1
Anger	2
Sadness	3
Disgust	4
Shame	5
Guilt	6

And an example of categorizing labels can be seen in TABLE IV.

TABLE IV.

Emotion in Datasets	*Categorical Array*
Joy	[1., 0., 0., 0., 0., 0., 0.]
Fear	[0., 1., 0., 0., 0., 0., 0.]
Anger	[0., 0., 1., 0., 0., 0., 0.]
Sadness	[0., 0., 0., 1., 0., 0., 0.]
Disgust	[0., 0., 0., 0., 1., 0., 0.]
Shame	[0., 0., 0., 0., 0., 1., 0.]
Guilt	[0., 0., 0., 0., 0., 0., 1.]

C. Word Embedding

Word embedding follows preparation. The process of mapping words from an existing dictionary to numerical vectors that contain real values is known as word embedding[2]. Pre-trained 300-dimensional word vectors from Wikipedia articles will be used. Although our dataset is very small, we may still use it to train the w2v model. However, the quality of the trained word vectors may not be as high as that of pretrained w2v.

D. LSTM Model

The next stage is to engage the LSTM Model procedure after word embedding. An algorithm for deep learning called LSTM was created using the RNN architecture[2]. The vanishing gradient problem of RNN can be resolved by LSTM using memory cells and gate units (input, forget, and output gates) to enable LSTM to read, store, and update data[2].

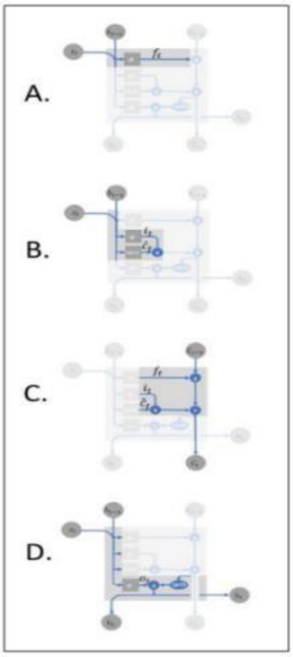

Fig. 1. Flowchart LSTM Model.

Recurrent neural networks (RNNs) with bidirectional LSTM (BiLSTM) can handle sequential data in both forward and backward orientations. We are actually utilizing a BiLSTM model, even though it was not indicated above. Compared to the regular LSTM, which can only process sequential data in one way, the BiLSTM can learn longer-range dependencies in sequential data since it can process the data in both forward and backward directions.

LSTM has a chain structure that contains four neural networks and different memory blocks called cells. This can be seen in fig. 2.

(X) Text

↓

Embedding (W2V pretrained in wiki)

↓

Deep Network (LSTM)

↓

Fully connected (Dense)

↓

Output Layer (SoftMax)

↓

Emotion class

Fig. 2. Processing Layers of LSTM Model[2].

Figure 3 illustrates the steps of the LSTM. A forget gate layer, with a sigmoid activation function that outputs 0 or 1, is the initial stage of the LSTM (Fig. 3(A)). 1 denotes "remember anything," and 0 indicates "let nothing through." Fig. 3(B) is used to determine which data will be saved next. The tanh layer generates new candidate values between -1 and 1, while the sigmoid layer (input gate layer) selects which value will be updated. The next phase, shown in fig. 3(C), involves multiplying the previous cell state by the forget gate's output in order to erase any information that is no longer needed and add the new data to the cell state. In Figure 3(D), the last action is to decide the output. To determine what portions of the cell state to output, we first run a sigmoid layer. After that, we multiply the cell state by the sigmoid gate's output after running it through tanh to force the values to be between -1 and 1, allowing us to output only the portions we choose to.

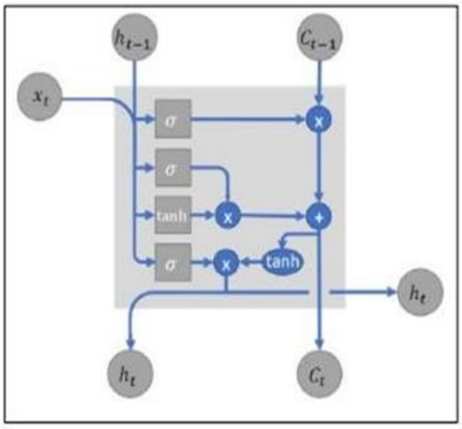

Fig. 3. LSTM Steps[8]

IV. RESULT AND DISCUSSION

Every epoch of the model was recorded during the LSTM training process as a checkpoint to get accuracy and loss progress. We have established 15 as the maximum epoch number. The LSTM's outputs for accuracy and loss are noted. The accuracy and loss outputs of the LSTM during the training phase are displayed in Figures 4 and 5. The LSTM model finished the 15 epoch with an accuracy of 71.22% and a loss of 0.8193.

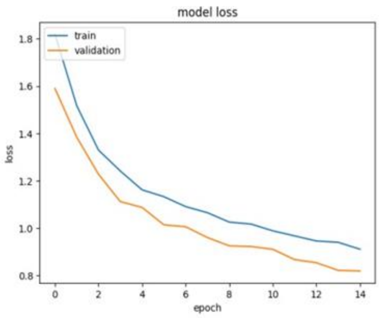

Fig. 4. Model Loss.

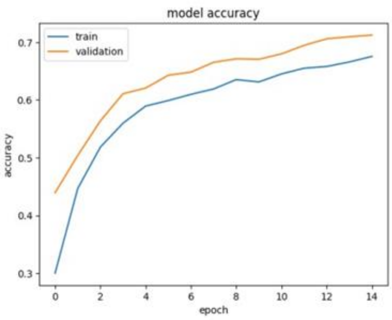

Fig. 5. Model Accuracy.

Fig. 6 shows a confusion matrix for the LSTM method. In this table, there are 7 emotions such as joy, fear, anger, sadness, disgust, shame, and guilt. From several tests we carried out, we found no errors in detecting emotions, although from the tests we carried out on this model it was found that the accuracy was only able to reach 71.22%. From this data it can be seen that grammar and word difficulty have an influence on the model we run, this also affects the output results of this model. One way that can be done to increase model accuracy is to run the model with larger training data, of course this requires more time. To determine the test's accuracy, precision, recall, and f1-score values, we used a confusion matrix. A classifier's confusion matrix can be used to analyze whether or not tuples from various classes are correctly recognized [2]. A confusion matrix provides a tabular layout of the different predictions and outcomes, which helps with the visualization of the results of a classification problem. It plots the actual and expected values for each classifier in a table.

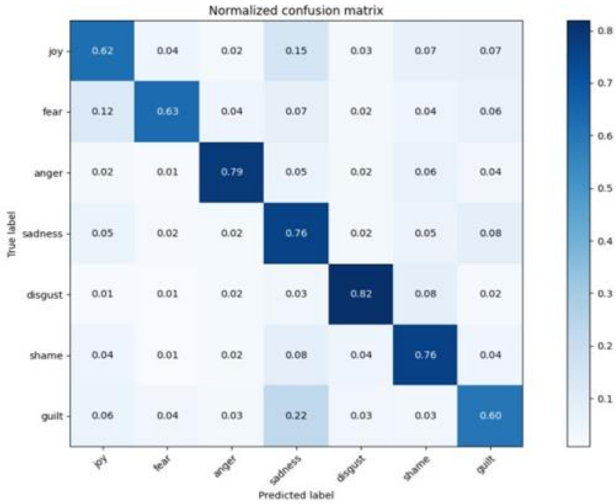

Fig. 6. Confusion Matrix.

V. CONCLUSION

In this paper, we presented our study about textual emotion detection utilizing an LSTM model to classify texts to different emotions. The model is trained to accurately classify 7 different emotions such as joy, fear, anger, sadness, disgust, shame, guilt. The model has given a 71.22% accuracy in doing the classification. Despite the seemingly low accuracy, the model has accurately guessed the emotion of texts we tested it 9 out of 10 times. The accuracy of this model is classify good because the majority of the text that goes through the model can be predicted well. In future research we hope to combining several models will provide better model accuracy result. The weakness of this model lies in the quantity and quality of the text in the dataset because the dataset that we ran on the model training only used 7000 data, so the accuracy result was lower than we imagined. Therefore, it is hoped to be able to combine other models such as the CNN model and more data will be used in future studies so that we can increase the accuracy of the models we develops.

REFERENCES

[1] J. L. Wu, Y. He, L. C. Yu, and K. Robert Lai, "Identifying Emotion Labels from Psychiatric Social Texts Using a Bi-Directional LSTM-CNN Model," IEEE Access, vol. 8, pp. 66638–66646, 2020, doi: 10.1109/ACCESS.2020.2985228.

[2] M. A. Riza and N. Charibaldi, "Emotion Detection in Twitter Social Media Using Long Short-Term Memory (LSTM) and Fast Text," International Journal of Artificial Intelligence & Robotics (IJAIR), vol. 3, no. 1, pp. 15–26, May 2021, doi: 10.25139/ijair.v3i1.3827.

[3] J. Singh, M. Wazid, D. P. Singh, and S. Pundir, "An embedded LSTM based scheme for depression detection and analysis," in Procedia Computer Science, Elsevier B.V., 2022, pp. 166–175. doi: 10.1016/j.procs.2022.12.019.

[4] S. Kusal, S. Patil, · Jyoti Choudrie, · Ketan Kotecha, · Deepali Vora, and I. Pappas, "A Review on Text-Based Emotion Detection-Techniques, Applications, Datasets, and Future Directions," 2022.

[5] S. K. Bharti et al., “Text-Based Emotion Recognition Using Deep Learning Approach,” Comput Intell Neurosci, vol. 2022, 2022, doi: 10.1155/2022/2645381.

[6] D. Kher and K. Passi, “Multi-label Emotion Classification using Machine Learning and Deep Learning Methods,” in International Conference on Web Information Systems and Technologies, WEBIST - Proceedings, Science and Technology Publications, Lda, 2022, pp. 128–135. doi: 10.5220/0011532400003318.

[7] A. Al 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,” IEEE Access, vol. 12, pp. 18416–18450, 2024, doi: 10.1109/ACCESS.2024.3356357.

[8] D. Haryadi and G. Putra Kusuma, “Emotion Detection in Text using Nested Long Short-Term Memory,” 2019. [Online]. Available: www.ijacsa.thesai.org

[9] N. Aslam, F. Rustam, E. Lee, P. B. Washington, and I. Ashraf, “Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model,” IEEE Access, vol. 10, pp. 39313–39324, 2022, doi: 10.1109/ACCESS.2022.3165621.

[10] Z. Wang, “Text Emotion Detection Based on Bi-LSTM Network,” Academic Journal of Computing & Information Science, vol. 3, pp. 129–137, 2019, doi: 10.25236/AJCIS.2020.030314.

[11] E. Batbaatar, M. Li, and K. H. Ryu, “Semantic-Emotion Neural Network for Emotion Recognition from Text,” IEEE Access, vol. 7, pp. 111866–111878, 2019, doi: 10.1109/ACCESS.2019.2934529.

[12] N. Chaudhari, R. Agarwal, M. Narwade, X. Yu, and J. Saniie, “Artificial Intelligence System for Emotion Recognition and Text Analytics,” 2022. [Online]. Available: http://ecasp.ece.iit.edu/

[13] M. J. Basha, S. Vijayakumar, J. Jayashankari, A. H. Alawadi, and P. Durdona, “Advancements in Natural Language Processing for Text Understanding,” in E3S Web of Conferences, EDP Sciences, Jul. 2023. doi: 10.1051/e3sconf/202339904031.

[14] S. Velampalli, C. Muniyappa, and A. Saxena, “Performance Evaluation of Sentiment Analysis on Text and Emoji Data Using End-to-End, Transfer Learning, Distributed and Explainable AI Models,” Journal of Advances in Information Technology, vol. 13, no. 2, pp. 167–172, Apr. 2022, doi: 10.12720/jait.13.2.167-172.

[15] D. Haryadi and G. Putra Kusuma, “Emotion Detection in Text using Nested Long Short-Term Memory,” 2019. [Online]. Available: www.ijacsa.thesai.org

[16] J. L. Wu, Y. He, L. C. Yu, and K. Robert Lai, “Identifying Emotion Labels from Psychiatric Social Texts Using a Bi-Directional LSTM-CNN Model,” IEEE Access, vol. 8, pp. 66638–66646, 2020, doi: 10.1109/ACCESS.2020.2985228.

[17] S. Ahmad, M. Z. Asghar, F. M. Alotaibi, and S. Khan, “Classification of Poetry Text into the Emotional States Using Deep Learning Technique,” *IEEE Access*, vol. 8, pp. 73865–73878, 2020, doi: 10.1109/ACCESS.2020.2987842.

[18] X. Lu, “Deep Learning Based Emotion Recognition and Visualization of Figural Representation,” *Front Psychol*, vol. 12, Jan. 2022, doi: 10.3389/fpsyg.2021.818833.

[19] L. Alzubaidi *et al.*, “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions,” *J Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00444-8.

[20] M. Spezialetti, G. Placidi, and S. Rossi, “Emotion Recognition for Human-Robot Interaction: Recent Advances and Future Perspectives,” *Frontiers in Robotics and AI*, vol. 7. Frontiers Media S.A., Dec. 21, 2020. doi: 10.3389/frobt.2020.532279.