SAKARYA UNIVERSITY JOURNAL OF COMPUTER AND INFORMATION SCIENCES

http://saucis.sakarya.edu.tr/

e-ISSN: 2636-8129
Publisher: Sakarya University

Vol. 7, No. 3, 2024, 338-345
DOI: 10.35377/saucis...1539409

RESEARCH ARTICLE

Emotion Recognition on Turkish Mobile Operator Turkcell's Call Center Calls

Yüksel Yurtay[1], Hüseyin Demirci[1*], Hüseyin Tiryaki[2], Tekin Altun[2]

[1] Sakarya University, Faculty of Computer and Information Sciences, Sakarya, Türkiye
[2] Turkcell Global Bilgi Marketing Consultancy and Call Center Services Inc., İstanbul, Türkiye

Corresponding author:
Hüseyin Demirci, Sakarya University,
Faculty of Computer and Information Sciences,
Sakarya, Türkiye,
huseyind@sakarya.edu.tr

Article History:
Received: 27.08.2024
Accepted: 30.09.2024
Published Online: 30.10.2024

ABSTRACT
A fundamental component of human intelligence is the capacity for feeling. In addition to being founded on logic and reason, human conduct is also greatly influenced by the emotions that people experience. For the purpose of this study, we classified one thousand real-life call center client voice data in the Turkish language based on the way they expressed their emotions using text emotion detection. We made use of Ekman's emotional labeling and techniques from the field of artificial intelligence, such as deep learning and other similar methods.

Keywords: Emotion recognition, Call center, Deep learning, Artificial intelligence

1. Introduction

Emotion is a complex feeling involving consciousness, physical sensation, and behavior that represents the individual meaning of an object, an occasion, or a situation [1]. An essential aspect of human intelligence is emotion. Human behavior is significantly impacted by emotion in addition to being based on logic and reason. Psychological research has identified three primary methods for modeling emotions [2,3].

The categorical method is grounded in the concept that there is a limited set of fundamental emotions that are universally acknowledged [2]. The prevailing paradigm utilized in research on emotion recognition is the one developed by Ekman [3], which encompasses six fundamental emotions: disgust, surprise, fear, anger, sadness and happiness.

The dimensional methodology claims that emotional states are interconnected and exhibit a systematic relationship with one another[2]. This method encompasses the diversity of emotions in three aspects [4,5]:

- Valence: This dimension pertains to the degree of positivity or negativity associated with an emotion.
- Arousal: This dimension pertains to the level of excitement or indifference associated with an emotion.
- Power: This dimension pertains to the extent of authority or control [2].

The appraisal-based technique can be viewed as an expansion of the dimensional technique. The system employs componential notions of emotion derived from appraisal concept [6]. According to this theory, an individual can feel an emotion when it is elicited by an assessment of circumstances, taking into account the individual's past experiences, objectives, and possibilities for action. Emotions are observed by examining alterations in various important elements, such as cognition, physiology, motivation, motor reactions, sensations, and expressions [2].

Human-computer interaction (HCI) has increased dramatically since human society entered the information age. Researchers from a wide range of subjects, including psychology, neurophysiology, cognitive science, computer science, and others, have been contributing to the study of emotional computing over the course of the past two decades. This technology has already been used in a number of different domains. The applications can be broadly classified into five categories: (1) Formal instruction. Robots can enhance the quality of teaching and learning by monitoring students' emotional states and focus in class. (2) Medical care. The medical robot incorporates cognitive and affective computing to aid doctors in treating psychiatric disorders and offers emotional solace to patients. (3) Sector of the economy that provides intangible goods or services to consumers. Robots employed in many sectors, such as banking, hospitals, catering, government services, and other businesses, can enhance consumers' experience by offering efficient and effective services across the full-service process. (4)

Cite as: Y. Yurtay, H. Demirci, H. Tiryaki, T. Altun, "Emotion Recognition on Turkish Mobile Operator Turkcell's Call Center Calls," *Sakarya University Journal of Computer and Information Sciences*, vol.7, no. 3, pp. 338-345, 2024. Doi: 10.35377/saucis…1539409

The entertainment sector. Integrating emotion recognition and interaction technologies into computer games can enhance the realism of virtual environments, alleviate player exhaustion, and enhance the overall enjoyment of the game. (5) Autonomous driving. The integration of emotion analysis and fatigue detection technologies in intelligent driving plays a crucial role in preventing traffic accidents.

Call centers play a crucial role in the realm of marketing and advertising. A call center is a collective of customer service specialists who manage telephone inquiries from prospective or current clients regarding a company's goods or products. Some call centers promote customer satisfaction and assistance, while others focus on sales expansion, lead generation, and customer acquisition. Call centers play a crucial role in creating a favorable client experience. Therefore, it is imperative that they continually deliver exceptional service in order to cultivate and maintain connections. Therefore, call center personnel must have a wide range of knowledge, display patience, and show helpfulness when interacting with customers. Given that call center personnel are typically the initial point of contact with consumers, it is crucial for them to effectively and appropriately respond to client emotions. Currently, it is advisable to apply emotion recognition algorithms to categorize customers based on how they are feeling and assign them to call center employees with varying levels of experience. This approach aims to enhance customer experience, optimize the response of call center employees, and improve overall company satisfaction.

In the present day, people are becoming more and more dependent on computers to carry out their day-to-day activities, which has resulted in an increased demand for the enhancement of human-computer interactions. The absence of fundamental knowledge hinders a computer's ability to identify and produce emotions. Consequently, extensive research has been carried out on the recognition of emotions. Emotion recognition can be broken down into three primary categories: emotion recognition based on facial expressions, emotion recognition based on voice, and emotion recognition based on text.

In this article, we analyze emotion through human speech. Since human speech is the main subject, we will not focus on human facial expressions regarding emotion recognition. Speech signals are the fastest and most basic means of human communication. Speeches are utilized as a rapid technique of establishing a connection between humans and computers. The speech signal contains valuable information that is not immediately apparent. Thus far, various forms of research have been conducted in the domain of recognition of speech. Despite significant progress in this domain, there remains a substantial disparity between the natural connection of computers and humans. The primary factor for this issue is the computer's incapacity to comprehend the user's emotions. Recognition of speech has emerged as a very complex and demanding area of study within speech processing, garnering significant interest from numerous experts in recent years. In addition, speech emotion recognition has the ability to extract significant meanings from speech, which in turn improves the performance of speech recognition systems [7–13].

In this paper, we studied speech-emotion recognition on call center speech data, which are real-life customer data provided by "Turkcell Global Bilgi Pazarlama Danışmanlık ve Çağrı Servisi Hizmetleri Inc." and handled the speech-emotion recognition in terms of text-based recognition. The common element of speech emotion recognition in literature is that most of the previous work is done by analyzing the voice characteristics of speech and speakers. In addition, the data sets used by researchers were generally recorded in a studio environment and by professional voice actors. However, in the real world, people express their emotions while speaking with more natural tones of voice and non-fluent speech. For example, a person who is frightened or confused should not be expected to speak correctly in terms of diction and grammar. This could cause the suggested models from earlier works to fail on real-life speech data. When we examine the literature based on emotion recognition in text, we can clearly see that most of the works are done on corpora and lexicons. These corpora and lexicons are very well written and labeled by professionals. Using these works in real-life speech data can also cause failure. In our work, we converted 1000 real-life speech data in the Turkish language, which are provided by "Turkcell Global Bilgi Pazarlama Danışmanlık ve Çağrı Servisi Hizmetleri Inc." and 5000 sample data in the English language into text conversation data and labeled it with Ekman's emotional labels. The novelty and significance of the work is that the data we used are purely natural human speech, which is collected via call center calls from customers of Turkish mobile operator Turkcell. This makes our work unique in emotion recognition based on voice and text areas. Also, the speech-to-text technique we use challenges emotion recognition based on voice and text areas, making our work valuable.

The rest of the paper is organized as follows. In Section 2, we mentioned text-based emotion recognition works in literature. Later on, in Section 3, the methodology of our study and dataset attributes are given in Section 3. Finally, results and discussions are mentioned in Section 4.

2. Background Work

Recognizing emotions in text, especially implicit emotion recognition, is a challenging job in Natural-Language-Processing (NLP) that necessitates Natural-Language-Understanding (NLU). Textual emotion recognition can be categorized into various levels: document, paragraph, phrase, and word. The challenge arises at the sentence-level, where emotions are conveyed through the semantics and interconnections of words; as the level progresses, the problem becomes more intricate. However, not all concepts are spoken with clarity; some may involve metaphors, sarcasm, or irony.

Various methodologies have been employed to identify emotions in written language. Studies have been conducted on methods that use keywords to accurately identify and classify explicit emotions. In order to recognize implicit emotions in text, a number of other ways have been specially presented. These include rule based methods [14,15], classical-learning-based methods [16,17], deep-learning methods [18,19], and hybrid methods [20,21]. In this paper, due to the challenging structure of our methodology, we used deep-learning algorithms to classify emotions in text; therefore, in the rest of the paper, we focus on deep-learning approaches.

2.1. Related Works on Deep Learning Approaches in Text Based Emotion Recognition

Deep learning is a subset of machine learning that involves algorithms acquiring knowledge from experience and comprehending the world based on a hierarchical structure of concepts, with each notion becoming progressively simpler. This methodology enables software to acquire intricate concepts by constructing them using foundational ones. The predominant deep learning model utilized in this context is Long-Short-Term-Memory (LSTM). The LSTM is a distinctive variant of Recurrent-Neural-Networks (RNN) that is proficient in managing long term dependencies. LSTM mitigates the issue of the disappearing or inflating gradient problem, which is frequently encountered with RNNs. Figure 1 delineates the primary stages of LSTM for the purpose of identifying emotions in text. Initially, the emotion dataset undergoes text preprocessing. The preprocessing procedures often involve tokenization, the elimination of stop-words, and lemmatization. Subsequently, the embedding-layer is constructed and then sent to single or several LSTM layers. Subsequently, the resulting data is inputted into a Densely-connected-Neural-Network (DNN) with the same number of units as the emotion tags, utilizing a sigmoid activation function to carry out the classification process.

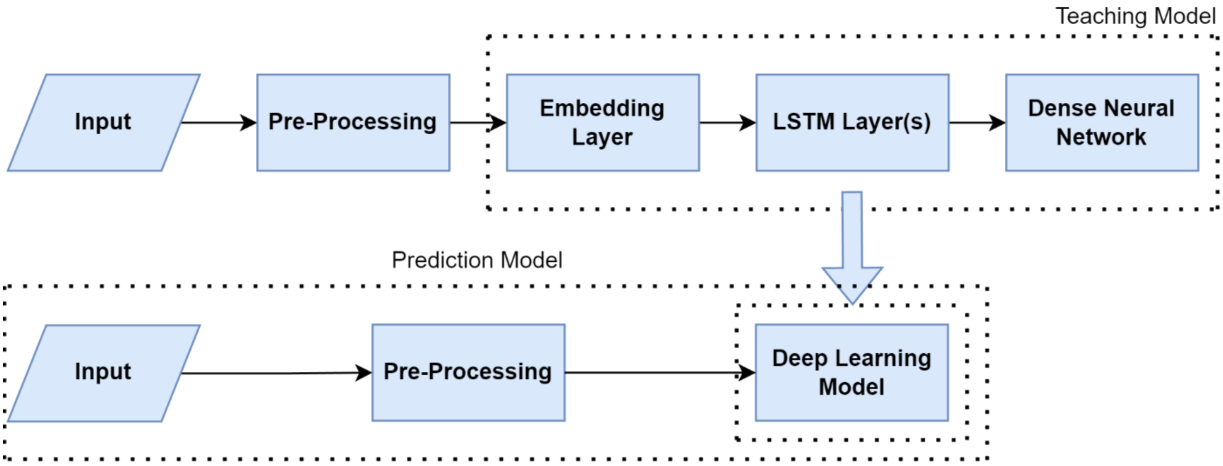

Figure 1. Main Steps of LSTM for Text Emotion Recognition

Shrivastava et al. [22] introduced a deep learning approach that utilizes the word2vec model to recognize emotions in multimedia text. The findings indicate that the accuracy of emotion labels anger and fear surpasses that of other emotion labels, although the recall and F1-score of pleasure outperform other emotion labels.

Rathnayaka et al. [23] introduced a deep-learning framework for detecting several emotions in microblogs. They employed an ekphrasis tool to prepare the data. In the work [24], they defined emotion recognition in the text as a binary classification task, employing two word embedding models: ConceptNet Numberbatch and fastText. The embedding layer was inputted into a Bidirectional-Gated-Recurrent-Unit (Bi-GRU) layer, and the pooling scores were inputted into a Deep Neural Network (DNN).

In the work [25], the authors introduced a deep learning model for identifying emotions in written conversations. They utilized three pre-trained embeddings, namely word2vec-twitter, GloVe, and ekphrasis. An embedding layer was inputted into a Bidirectional Long Short-Term Memory (Bi-LSTM) layer, which was then came next by an attention layer and a Convolutional Neural Network (CNN) layer. The results of the Bi-LSTM and CNN layers were combined, and global max pooling was performed. The scores of pooling were inputted into a Deep Neural Network (DNN) that utilized a softmax activation function to perform classification.

In the work [26], they introduced a deep learning model for identifying emotions in written conversations. They replaced emojis with appropriate emotion terms and inputted the embeddings into a Bi-LSTM layer. The output of the Bi-LSTM was subjected to an inner product operation with the attention weights, and the result was then sent into another Bi-LSTM layer. The scores of pooling were inputted into an LSTM layer and subsequently passed through a DNN with a softmax activation function.

In article [27], researchers introduced a deep-learning algorithm for identifying emotions in written conversations.. The model involves combining three parts of the dialogue into an embedding layer. The output is then passed through three subsequent layers of Bi-LSTM, which are trained using average stochastic gradient descent. Self-attention technique is utilized, then average pooling is performed. The distinction between the two pooled scores is used as input to a two-layer deep neural network (DNN), which is followed by a softmax function to obtain emotion labels. The language model was trained using the Wikitext-103 dataset.

Xiao [28] presented a deep-learning framework for detecting emotions in written conversations. The model utilized an ekphrasis tool to preprocess the text. The researchers optimized multiple models, such as the Universal-Language-Model (ULM), the BERT model, OpenAI's Generative-PreTraining (GPT) model, DeepMoji model, and a DeepMoji model trained via NTUA embedding. The ULM model exhibited superior performance, whilst the DeepMoji model, trained via NTUA embedding, achieved the second-highest ranking. Nevertheless, combining these models yielded the highest outcome.

In the work [29], they introduced a deep learning framework for detecting emotions in written conversations. The framework has four submodels: a three input submodel (INP3), two output submodels (OUT2), a sentence encoder submodel, and a Bidirectional-Encoder-Representation from Transformers (BERT). The results demonstrated the characteristics acquired by the INP3 and OUT2 componenets yielded superior efficiency compared to the characteristics acquired by the USE and BERT componenets. Nevertheless, the combination of the four componenets yielded the most optimal performance outcome using SVM-n.

3. Material and Methods

3.1 Methodology

This paper outlines two primary methodologies for identifying emotions from voice and textual data. After the data collected from the transcription of audio recordings into text, a Text-Based Classification model is utilized in order to classify the various emotional states. In the second step of the process, a Voice-Based Classification model is created in order to estimate emotional states by directly extracting characteristics from recordings. Furthermore, its objective is to conduct a more extensive analysis by merging the outcomes of these two approaches. This study conducts a comparison of the three strategies and specifically aims to ascertain the most efficacious way.

For the voice model training, English voice data were used, where features such as understanding emotional expressions, stress, and innuendo were similar. This approach aims to enable the model to better understand the tonal characteristics and accents of the language. English audio data provides a basis for learning universal features of the language and integrating this information with text data from different languages. During this phase of the study, we utilized Turkish speech data obtained from Turkcell. However, the quantity of data was insufficient for training and testing the model. To compensate for this, we incorporated sample speech recordings in English to identify universal speech characteristics that can be used to detect emotions in human speech.

In the text model training, completely Turkish text data is used to learn the structural and semantic features of the language. In this way, while the model learns the essential elements of the language, such as grammar rules, word usage, and sentence structure, it also masters the cultural and local features of the language. Turkish text data is critical to ensure that the model better understands the semantic layer of the language and interprets it correctly. The text-based classification technique is specifically designed to perform sentiment classification on textual data obtained from audio sources. This model consists of a series of deep learning layers specifically designed to understand and analyze the features of textual content. Every layer is specifically developed with a unique purpose to optimize the effectiveness and accuracy of the model. These layers form of four components: the Embedding Layer, Bidirectional GRU Layer, Attention Mechanism, and Flatten and Dense Layers.

After both models underwent specialized training in their respective fields, the results obtained using voice and text data were compared and analyzed. Thanks to this integration, the model aims to understand better and interpret both the tonal characteristics of the voice and the structural and semantic elements of the voice. As a result, this bidirectional training approach plays a crucial function in enhancing accuracy. and effectiveness of our language model.

The audio data we used consists of the following attributes;

- Sampling Rate: 8 kHz
- Bit Depth: 8 bits
- Data Labels: Positive, Negative, Neutral
- Distribution:
- Initial data set: Even distribution for Positive, Negative, and Neutral
- 5000 sample records in English and 1000 records in Turkish

We used %70 of the data for training, %15 of the data for validation, and %15 for the data for testing. Each sample record consists of multiple sentences that can express different emotional states in conversations. For instance, a dialogue can commence with favorable sentiments. As the talk unfolds, the client may experience anger or frustration, leading to the characterization of the interaction as negative. In this example, during the initial phases of the conversations, the sentences obtained from the call may be categorized as positive, but in the later stages of the conversation, the extracted sentences can be classified as negative. As a result, this conversation can be labeled negative for a client. Nevertheless, this interaction might be classified as neutral since the operator fulfilled their duty by providing the client with the most optimal answer. Based on this frame, it is evident that the Turkish recordings dataset consists of around 3000 sentences, equivalent to 1000 recordings.

In this study, different approaches were adopted to process and analyze voice and text-based data. While traditional machine learning algorithms are used on audio data, deep learning models are preferred for text-based data. The Voice-Based classification approach utilizes Mel-frequency cepstral coefficients (MFCC) features extracted from audio recordings to train various machine learning models. The models include Support Vector Machines (SVM), Support Vector Regression (SVR), and Logistic Regression. We used Librosa and pydub libraries for voice analysis and processing. Using these libraries, we processed the audio files at an 8 kHz sample rate. These models are preferred for classification and regression analysis of audio data.

Deep learning models have been developed using LSTM and GRU-based models, dropout and normalization layers, Adam optimizer, and unique learning rate settings to process text-based data. These structures are designed to learn long-term dependencies and relationships in text data. While TensorFlow is used in model training and development, PyTorch support is also provided. Thanks to GPU support, TensorFlow offers a powerful platform, especially for training deep learning models.

Machine learning and feature extraction models for audio data and deep learning techniques for text data were optimized to meet the different needs of this study.

To assess the model, a precise matching criterion was employed to analyze three distinct types of outcomes. False negative (FN) and false positive (FP) refer to inaccurate negative and positive predictions, respectively. True positives (TP) refer to accurate positive predictions that align with the actual correct predictions. The assessment is founded on the performance metrics precision (P), recall (R), and F-score (F1). Recall refers to the proportion of positive cases that are correctly identified as positive, and it is calculated by dividing the number of correctly labeled positive results by the total number of positive cases. Equations for calculating precision (P), recall (R), and F-score (F1) are given in Equation 1, Equation 2, and Equation 3, respectively.

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

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

$$\mathrm{F1} = \frac{2 * P * R}{P + R} \tag{3}$$

4. Results and Discussion

The model developed in this study is an essential step for sentiment analysis in the Turkish language. The results obtained in the initial stages demonstrate the model's capacity to learn basic structures and its potential to classify emotional expressions. Ongoing studies aim to improve performance on the generalized data set and enable the model to be used more effectively in real-world scenarios.

Since we aim to learn the emotional state of the human being via audio data, which can be obtained from a phone call with deep learning methods, we only focused on positive and negative emotional states and ignored neutral states in the results. However, we trained our model to distinguish all the emotional states: positive and negative. The prediction results of the model are given in Table 1.

Table 1. Predicted Result of Both English and Turkish Data

	Actual Positive	**Actual Negative**
Predicted Positive	396	104
Predicted Negative	36	464

Table 2. Statistical Analysis of Both English and Turkish Data

	Positive Emotions	**Negative Emotions**
Recall	0.915662651	0.816513761
Precision	0.791666667	0.927083333
F1	0.849162011	0.868292683

As we can see in Table 2, our model has achieved a recall value of 0.915662651 in positive emotions and 0.816513761 in negative emotions. Also, the precision value of positive emotions is 0.791666667, and the value of negative emotions is 0.927083333. We can calculate the F1 scores of positive and negative emotions with these values: positive emotions=0.849162011 and negative emotions=0.868292683.

The statistical analysis and the results show us our model can distinguish positive and negative emotions from each other with a slight margin of error. The data we used in our model are real-world data gathered from call center conversations, which have never been used in literature and are purely human. The originality of the work is that we used actual human voice data. In other related works, all recordings are made in a studio and are not natural conversations between two humans. Also, we used a hybrid model that decodes speech into text and uses this text data to analyze a semantical emotion in the conversation context.

5. Conclusion

As a result, this document provides a research article that focuses on the identification of emotions in call center conversations involving the Turkish mobile operator Turkcell. The study employs text emotion detection techniques and artificial intelligence methods, specifically deep learning, to categorize real-life call center client voice data according to their expressed emotions. The findings demonstrate that the model exhibits a high level of accuracy in discerning between positive and negative emotions, with a minimal degree of error. This research is a joint effort between Sakarya University and Turkcell, with the goal of enhancing the model and expanding the dataset for future emotion recognition.

This work will provide assistance to both our collaborator Turkcell as well as the marketing and customer service departments of businesses. In our future works, we will exclude the English voice analysis and augment the dataset exclusively with Turkish call center recordings. We will improve the training of the model by incorporating the expanded dataset to capture the structural and phonetic characteristics of the Turkish language. During the final stage, we will conduct simultaneous testing of the trained model on live call center calls.

This work is an early work of a project between Sakarya University and "Turkcell Global Bilgi Pazarlama Danışmanlık ve Çağrı Servisi Hizmetleri" Incorporation.

References

[1] R. C. Solomon, "Emotion Definition, Examples, Scope, Structures, & Facts Britannica." Jan. 2024. Accessed: Jan. 17, 2024. [Online]. Available: https://www.britannica.com/science/emotion

[2] D. Grandjean, D. Sander, and K. R. Scherer, "Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization," *Conscious. Cogn.*, vol. 17, no. 2, pp. 484–495, Jun. 2008, doi: 10.1016/j.concog.2008.03.019.

[3] P. Ekman, "Basic Emotions," in *Handbook of Cognition and Emotion*, John Wiley & Sons, Ltd, 1999, pp. 45–60. doi: 10.1002/0470013494.ch3.

[4] H. Binali and V. Potdar, "Emotion detection state of the art," in *Proceedings of the CUBE International Information Technology Conference*, in CUBE '12. New York, NY, USA: Association for Computing Machinery, Sep. 2012, pp. 501–507. doi: 10.1145/2381716.2381812.

[5] X. Jin and Z. Wang, "An Emotion Space Model for Recognition of Emotions in Spoken Chinese," in *Affective Computing and Intelligent Interaction*, J. Tao, T. Tan, and R. W. Picard, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2005, pp. 397–402. doi: 10.1007/11573548_51.

[6] K. R. Scherer, "Appraisal theory," in *Handbook of cognition and emotion*, Hoboken, NJ, US: John Wiley & Sons Ltd, 1999, pp. 637–663. doi: 10.1002/0470013494.ch30.

[7] J. Nicholson, K. Takahashi, and R. Nakatsu, "Emotion Recognition in Speech Using Neural Networks," *Neural Comput. Appl.*, vol. 9, no. 4, pp. 290–296, Dec. 2000, doi: 10.1007/s005210070006.

[8] A. Mikuckas, I. Mikuckiene, A. Venckauskas, E. Kazanavicius, R. Lukas, and I. Plauska, “Emotion Recognition in Human Computer Interaction Systems,” *Elektron. Ir Elektrotechnika*, vol. 20, no. 10, pp. 51–56, Dec. 2014, doi: 10.5755/j01.eee.20.10.8878.

[9] W.-J. Yoon, Y.-H. Cho, and K.-S. Park, "A Study of Speech Emotion Recognition and Its Application to Mobile Services," in *Ubiquitous Intelligence and Computing*, J. Indulska, J. Ma, L. T. Yang, T. Ungerer, and J. Cao, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2007, pp. 758–766. doi: 10.1007/978-3-540-73549-6_74.

[10] D. J. France, R. G. Shiavi, S. Silverman, M. Silverman, and M. Wilkes, "Acoustical properties of speech as indicators of depression and suicidal risk," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 7, pp. 829–837, Jul. 2000, doi: 10.1109/10.846676.

[11] T. H. Falk and W.-Y. Chan, "Modulation Spectral Features for Robust Far-Field Speaker Identification," *IEEE Trans. Audio Speech Lang. Process.*, vol. 18, no. 1, pp. 90–100, Jan. 2010, doi: 10.1109/TASL.2009.2023679.

[12] S. Patil and G. K. Kharate, "A Review on Emotional Speech Recognition: Resources, Features, and Classifiers," in *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, Greater Noida, India: IEEE, Oct. 2020, pp. 669–674. doi: 10.1109/ICCCA49541.2020.9250765.

[13] H. Gunes and M. Pantic, "Automatic, Dimensional and Continuous Emotion Recognition," *Int. J. Synth. Emot. IJSE*, vol. 1, no. 1, pp. 68–99, Jan. 2010, doi: 10.4018/jse.2010101605.

[14] S. Y. M. Lee, Y. Chen, and C.-R. Huang, "A Text-driven Rule-based System for Emotion Cause Detection," in *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, D. Inkpen and C. Strapparava, Eds., Los Angeles, CA: Association for Computational Linguistics, Jun. 2010, pp. 45–53. Accessed: Jan. 17, 2024. [Online]. Available: https://aclanthology.org/W10-0206

[15] O. Udochukwu and Y. He, "A Rule-Based Approach to Implicit Emotion Detection in Text," in *Natural Language Processing and Information Systems*, C. Biemann, S. Handschuh, A. Freitas, F. Meziane, and E. Métais, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 197–203. doi: 10.1007/978-3-319-19581-0_17.

[16] C. O. Alm, D. Roth, and R. Sproat, "Emotions from Text: Machine Learning for Text-based Emotion Prediction," in *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, R. Mooney, C. Brew, L.-F. Chien, and K. Kirchhoff, Eds., Vancouver, British Columbia, Canada: Association for Computational Linguistics, Oct. 2005, pp. 579–586. Accessed: Jan. 17, 2024. [Online]. Available: https://aclanthology.org/H05-1073

[17] S. Aman and S. Szpakowicz, "Identifying Expressions of Emotion in Text," in *Text, Speech and Dialogue*, V. Matoušek and P. Mautner, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2007, pp. 196–205. doi: 10.1007/978-3-540-74628-7_27.

[18] C. Baziotis *et al.*, "NTUA-SLP at SemEval-2018 Task 1: Predicting Affective Content in Tweets with Deep Attentive RNNs and Transfer Learning," in *Proceedings of the 12th International Workshop on Semantic Evaluation*, M. Apidianaki, S. M. Mohammad, J. May, E. Shutova, S. Bethard, and M. Carpuat, Eds., New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 245–255. doi: 10.18653/v1/S18-1037.

[19] A. Ezen-Can and E. F. Can, "RNN for Affects at SemEval-2018 Task 1: Formulating Affect Identification as a Binary Classification Problem," in *Proceedings of the 12th International Workshop on Semantic Evaluation*, M. Apidianaki, S. M. Mohammad, J. May, E. Shutova, S. Bethard, and M. Carpuat, Eds., New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 162–166. doi: 10.18653/v1/S18-1023.

[20] W. Amelia and N. U. Maulidevi, "Dominant emotion recognition in short story using keyword spotting technique and learning-based method," in *2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA)*, Aug. 2016, pp. 1–6. doi: 10.1109/ICAICTA.2016.7803131.

[21] S. Gievska, K. Koroveshovski, and T. Chavdarova, "A Hybrid Approach for Emotion Detection in Support of Affective Interaction," in *2014 IEEE International Conference on Data Mining Workshop*, Dec. 2014, pp. 352–359. doi: 10.1109/ICDMW.2014.130.

[22] K. Shrivastava, S. Kumar, and D. K. Jain, "An effective approach for emotion detection in multimedia text data using sequence based convolutional neural network," *Multimed. Tools Appl.*, vol. 78, no. 20, pp. 29607–29639, Oct. 2019, doi: 10.1007/s11042-019-07813-9.

[23] P. Rathnayaka *et al.*, "Gated recurrent neural network approach for multilabel emotion detection in microblogs," *arXiv.org*. Jul. 2019. Accessed: Jan. 20, 2024. [Online]. Available: https://arxiv.org/abs/1907.07653v1

[24] A. Seyeditabari, N. Tabari, S. Gholizadeh, and W. Zadrozny, "Emotion detection in text: focusing on latent representation," *arXiv.org*. Jul. 2019. Accessed: Jan. 20, 2024. [Online]. Available: https://arxiv.org/abs/1907.09369v1

[25] S. Ge, T. Qi, C. Wu, and Y. Huang, "Thu_ngn at semeval-2019 task 3: dialog emotion classification using attentional lstm-cnn," in *Proceedings of the 13th International Workshop on Semantic Evaluation*, Minneapolis, Minnesota, USA: Association for Computational Linguistics, 2019, pp. 340–344. doi: 10.18653/v1/S19-2059.

[26] L. Ma, L. Zhang, W. Ye, and W. Hu, "PKUSE at SemEval-2019 Task 3: Emotion Detection with Emotion-Oriented Neural Attention Network," in *Proceedings of the 13th International Workshop on Semantic Evaluation*, Minneapolis, Minnesota, USA: Association for Computational Linguistics, 2019, pp. 287–291. doi: 10.18653/v1/S19-2049.

[27] W. Ragheb, J. Azé, S. Bringay, and M. Servajean, "LIRMM-Advanse at SemEval-2019 Task 3: Attentive Conversation Modeling for Emotion Detection and Classification," in *Proceedings of the 13th International Workshop on Semantic Evaluation*, Minneapolis, Minnesota, USA: Association for Computational Linguistics, 2019, pp. 251–255. doi: 10.18653/v1/S19-2042.

[28] J. Xiao, "Figure Eight at SemEval-2019 Task 3: Ensemble of Transfer Learning Methods for Contextual Emotion Detection," in *Proceedings of the 13th International Workshop on Semantic Evaluation*, Minneapolis, Minnesota,

USA: Association for Computational Linguistics, 2019, pp. 220–224. doi: 10.18653/v1/S19-2036.
[29] A. Basile, M. Franco-Salvador, N. Pawar, S. Štajner, M. Chinea Rios, and Y. Benajiba, "SymantoResearch at SemEval-2019 Task 3: Combined Neural Models for Emotion Classification in Human-Chatbot Conversations," in *Proceedings of the 13th International Workshop on Semantic Evaluation*, Minneapolis, Minnesota, USA: Association for Computational Linguistics, 2019, pp. 330–334. doi: 10.18653/v1/S19-2057.

Article Information Form

Authors Contributions: All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

Conflict of Interest Notice
Authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Ethical Approval
It is declared that during the preparation process of this study, scientific and ethical principles were followed, and all the studies benefitted from are stated in the bibliography.

Plagiarism Statement: This article has been scanned by iThenticate ™.