DAN: a Deep Broad Learning Technique for Emotion Classification in Textual / Speech Conversations

1 P.Madhavi Latha, Assistant Professor, Dept of Information technology , Malla Reddy Engineering College and Management Sciences, Hyderabad

2 G.Kasi Reddy, Assistant Professor, Dept of Computer Science and Engineering, Guru Nanak Institute of Technology, Hyderabad

3 V.Pavani, Assistant Professor, Dept of Computer Science and Engineering, TKRCET, Hyderabad

4 Nunavath Jagender, Dept of ET, Assistant Professor, GNIT (GURU NANAK Institute of Technology), Hyderabad

Abstract – Sentiment analysis is the process of gauging people's sentiments, attitudes, and emotions towards specific targets, such as activities, organizations, services, topics, and products. Emotion recognition, a subset of sentiment analysis, predicts distinct emotions rather than simply categorizing sentiments as positive, negative, or neutral. Recent research has delved into language and facial expressions to discern emotions. Combining Convolutional Neural Network (CNN) with a proposed algorithm enhances classification accuracy and reduces processing time by addressing gradient saturation issues and mitigating data imbalance effects. This approach is demonstrated in a binary sentiment problem context. Various methods have been proposed to identify emotions from text using natural language processing (NLP) techniques, including keyword-based, dictionary-based, and machine learning approaches. However, keyword- and vocabulary-based methods are constrained by semantic relationships and have limitations. In this study, we introduce a hybrid model (combining machine learning with deep learning) for emotion identification in text. On the other hand, in the deep learning realm, the LSTM model achieves the highest accuracy at 83%, while the CNN model attains the highest F1 score of 72.39%. Our hybrid model delivers a precision rate of 74.2%, recall rate of 78.63%, an F1 score of 75.01 and an overall precision rate of 74.2%.

Keywords: Convolutional Neural Networks; Emotion Classification; Text mining, Deep learning and Sentiment analysis

1. Introduction

Emotion recognition capabilities are essential not only for successful interpersonal relationships but also for humanmachine interaction. Understanding and knowing how to react to emotions significantly improves the interaction and its outcome. It is therefore a crucial component in the development of empathetic machines, which substantially enriches the experiences these can provide. Emotion Recognition in Conversation (ERC) modules are useful for a wide range of applications, from automatic opinion mining, to emotion-aware conversational agents and as assisting modules for therapeutic practices. ERC is therefore an actively growing research field, with high applicability and potential.

It also provides comprehensive descriptions of Deep Learning methods from the Multi-Layer Perceptron, Recurrent Neural Networks, Long Short-Term Memory Networks, Gated Recurrent Units, Convolutional Neural Networks, Graph Neural Networks, Attention Mechanisms to the Transformer, in the light of Emotion Recognition in Conversations. While these methods are transversal to all emotion recognition tasks, and therefore all emotion recognition surveys, some are more appropriate for ERC. Examples are Transformer-based architectures that are better at preserving long-term dependencies between sequences of utterances than recurrent neural networks. Related relevant surveys, but not specific for conversations, covered sentiment analysis research challenges and directions [6] and textual emotion recognition and its challenges [17], the latter including a section on textual emotion recognition in dialogue, elaborating on utterance context modelling and dynamic emotion modeling.

In recent times, researchers have proposed various methods to detect the emotions of the text, such as keyword-based, lexical affinity, learning-based, and hybrid models [3]. In the beginning, they introduced a rule-based approach that consisted of two approaches, namely, lexical affinity-based and keyword-based. Later on, a new approach came into existence, i.e., the learning-based approach. This method was more accurate and gave better results. In a learning-based approach, different models are used to detect emotion. Many researchers have also started combining the approaches and making them hybrid in the search for high accuracy. As per the study, deep learning models show better accuracy than machine learning models for large sizes of text or data. But for small data, machine learning gives us better accuracy. Still, none of the approaches gave a complete solution to detect the emotion from a given text.

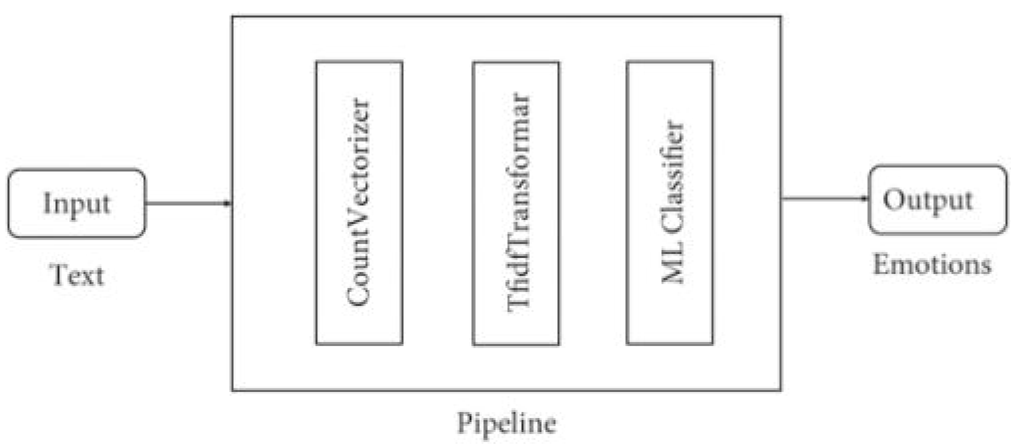

Fig 1: Basic process of Emotion Classification

There were many limitations in the existing solutions, such as that they did not have a list of all the emotions. The existing lists have an inadequate vocabulary of words in the lexicon, disregarded words, semantics-based context, low extractions of contextual information from the given sentences, do not perform well for detecting some specific emotions; weak context information extraction, loose semantic feature extraction, less computational speed, ignored relations between features, an inadequate amount of data, and a high number of misclassifications. Some models were not suited well for frequently occurring emojis, weak semantic information extraction, and the structure of the sentence. It differs from model to model. There were many limitations in this system that were fulfilled by previous researchers. The proposed model has fulfilled many of the existing limitations.

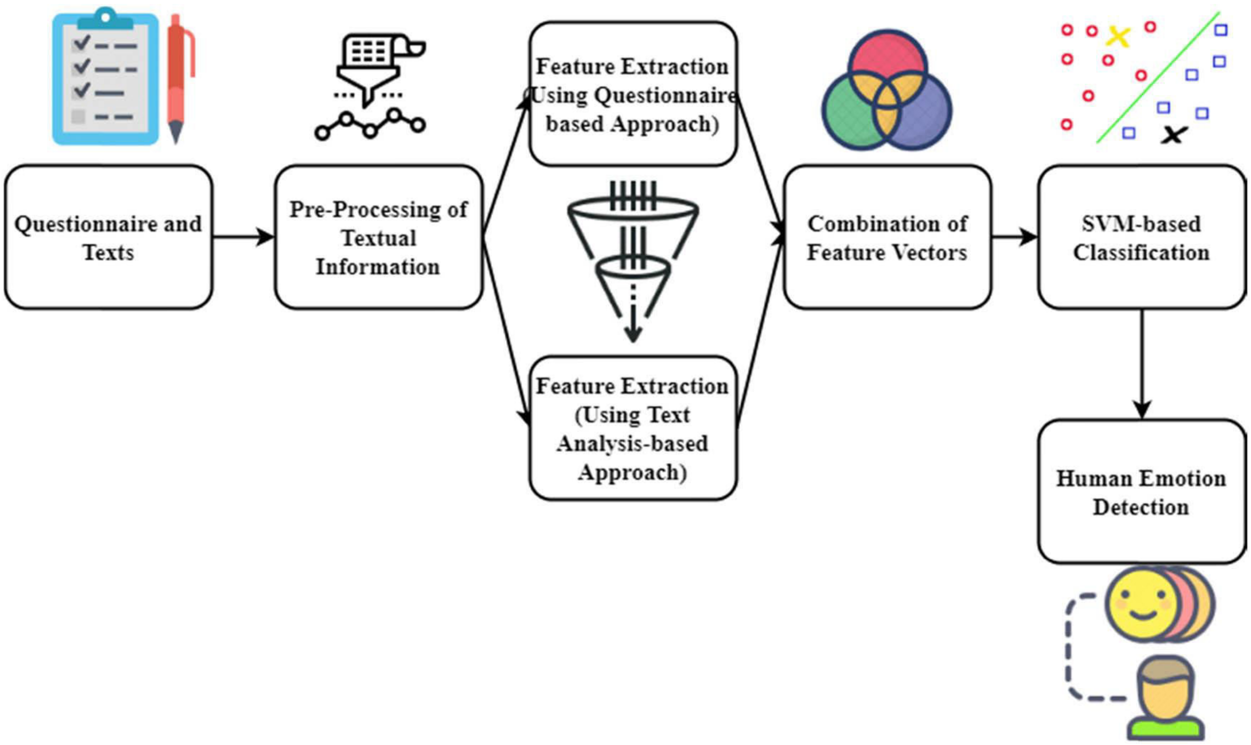

Fig 2: classifying text to generate emotion

Emotion detection is one of the big advantages of human-machine interaction as a nonliving thing can sense or feel like a human being. Our proposed model can detect emotions from text sentences that are tasteless as they do not have any tone or expression. Many researchers have worked on a single dataset. But we have worked on three datasets which include the textual form of simple sentences, tweets, and dialogs to detect emotions. Our text-based emotion recognition model can be implemented on any system. For business potential, this model can help to find emotions from customer reviews, services, give security for social media users, and many others.

2. Literature Survey

Seal et al. [4] have performed emotion detection with a keyword-based approach mainly focused on phrasal verbs. They used ISEAR [5] data, preprocessed the data, and then applied the keyword-based approach. They discovered several phrasal verbs that should have been associated with emotion terms but were not, and so they built their own database. They recognized phrasal verbs and keywords synonymous with various emotions and categorized them using their database. They did, however, achieve a much higher accuracy of 65%, but they were unable to address the researcher's existing issues, such as an insufficient list of emotion keywords and a lack of respect for word semantics in meaning.

F. A. Acheampong [20] surveyed the concept of emotion detection (ED) from texts and highlighted the main approaches adopted by researchers in the design of text-based ED systems. Navarrete Verma [16] P. Nandwani and R. Verma [21] described the process used to create an emotion lexicon enriched with the emotional intensity of words and focused on improving the emotion analysis process in texts [13]. Sailunaza and Alhajj [17] K. Sailunaz and R. Alhajj [22] used Twitter data to detect emotion and sentiment from text. They exploited sentiment and emotion scores to generate generalized and personalized recommendations for users based on their Twitter activity [4].

To the best of our knowledge, the only survey on Emotion Recognition in Conversations is the one by Poria et al., from 2019 [89]. It focuses on the challenges and advances in text ERC, pointing out conversational context modelling, speaker specific modelling, the presence of emotion shift, multiparty conversations and the presence of sarcasm. Since 2019, a plethora of works resorting to novel deep learning architectures that followed, especially top performing gated and graph-based neural networks and transformer-based language models, were used to tackle those aspects. This survey describes how state-of-the-art works addressed those challenges. Furthermore, this survey discusses additional additional challenges for this task, namely the ones pertaining to real world ERC such as real time emotion recognition, recognizing emotion causes, different taxonomies across datasets, multilingual ERC, and interpretability.

Rodriguez et al. [13] use emotion analysis to identify hate speech on social media. Their aim with this research was to locate and analyse the unstructured data of selected social media posts that intend to spread hate in the comment sections. Cao et al. [14] exploited machine and deep learning approaches to evaluate emotion in textual data. They also highlight the issues and challenges regarding emotion detection in text. Acheampong et al. [15]

Xu et al. [8] P. Xu et al. [12] has proposed an Emo2Vec method that encodes emotional semantics into vector form. They have trained Emo2Vec on a multitask learning framework by using smaller and larger datasets (smaller datasets such as ISEAR, WASSA, and Olympic). It shows that their results are better than those of Convolution Neural Network (CNN), DeepMoji embedding, and more. They have utilized their work on emotion analysis, sarcasm classification, stress detection, etc. Finally, the model Emo2Vec, when combined with Logistic Regression and GloVe, can achieve more competitive results.

Aagheb et al. [9] W. Ragheb et al. [13]worked on detecting emotions from textual conversations through the help of learning-based model. Their data comprises 6 types of emotions that Paul Ekman has [1] described. In their methods, two phases of encoder and classification are present. After the data is collected, it is tokenized and passed to an encoder, which then passes it on to Bi-LSTM units that have been trained using average stochastic gradient descent (ASGD). To avoid over-fitting, they have applied dropouts between the LSTM units. Then, to focus on specific emotion-carrying conversations, a self-attention mechanism was used. The data was classified into its respective categories through the help of a dense layer and a SoftMax activation. The model showed an F1 score of 75.82%.

The challenge of identifying the emotional qualities of voice, regardless of the semantic meaning, is known as speech emotion recognition (SER). While people are capable of performing this activity efficiently as a natural aspect of voice communication, the capacity to do so autonomously through programmed technologies is indeed a work in progress. As it offers perspective on human mental processes, emotion identification from speech signals is a frequently investigated topic in the construction of human–computer interface (HCI) models. In HCI, it is frequently necessary to determine the emotion of persons as mental feedback. An attempt is made in this study to distinguish seven different emotions using speech signals: sadness, anger, disgusted, pleased, surprised, enjoyable, and neutrality mood. For the identification of emotion, the suggested method uses a signals preprocessing method based on the randomness measure. The signals are first normalized to reduce noise. Due to the obvious changing length and continual form of voice signals, emotions identification requires both locally and globally information. Local features depict dynamic behavior, while feature points reveal statistic factors such as standard error, median, and lowest and maximum values. The SER system includes several features, including spectrum characteristics, sound quality characteristics, and Teager energy operator-based characteristics. Prosodic features are those that are based on the human perception, such as rhythm and inflection. These characteristics are based on three factors: power, length, and frequency response. From of the heavily processed signals, a features vector is generated that evaluates the random feature for all of the emotional responses. Then, using mutual information (MI), the feature vector is utilized to choose from the entire set. The feature vectors are then categorized using the BOAT method and association rule mining. Experiments were carried out on the TESS dataset for several metrics, and the performance of the suggested method outperformed the state-of-the-art methods.

3. Emotion Classification Techniques

3.1 Emotion classification based on CNN :

CNN is frequently selected for emotion classification and have achieved excellent results in a range of studies. Stojanovski et al. [14] proposed a model to improve sentiment analysis emotion classification. They proposed a model that combined CNN with GloVe word embedding to obtain enhanced accuracy. They included a sigmoid activated layer and a tangent activated layer in the CNN to improve performance further. They also suggested that adding more layers to the network would improve performance. This research achieved considerable improvements with an accuracy of 58.84 in emotion identification and an F1 score of 64.88 in sentiment analysis. However, the training data imbalance problem still needs to be solved [4]. Zhang et al. [9] improved sentiment classification in a text by proposing a model using three different word embeddings to integrate the sentiment meaning into semantic embeddings. The coordination of the attention vector with CNN enables the model to extract both local and global features.

3.2 Emotion classification based on Long Short-Term Memory (LSTM):

Unlike CNN, LSTM, a variation of Recurrent Neural Networks (RNN) [10], can identify long-term dependencies in the input [11, 12]. Different models using LSTM have been proposed to improve the capability of emotion classification. Mahmoudi et al. [30] compared a range of approaches from traditional machine learning methods to deep learning approaches for improved emotion prediction in the stock market. Then they introduced a solution using LSTM to seize long-term dependencies to emphasize the linguistic objects. They adopted domain-specific word embedding GloVeST with non-static capabilities that allow manipulation of word positions to reflect the word's positive or negative character more accurately. The solution achieved enhanced investor emotion classification performance, as confirmed by the Wilcoxon Sum-Rank Test (WSuRT) with $p \leq 0.0283$. However, the researchers identified the limits of their work as originating from the small size of the data set as they had to rely on 'StockTwits' alone. Furthermore, this research classification is limited to only two types (i.e. bullish and bearish) which becomes a problem in a bullish market where the training data class imbalance is of such magnitude that it requires extensive modifications before it can be applied to other domains [6]

3.3 Random forest:

Random Forest is also referred to as a random decision forest . It is a method based on a decision tree and has been used widely for classification as well as regression. Introduced by Ho in 1995 [7] but was improved further by Breiman in 2001 [9] and published a version that is modified and being used currently. The random forest starts with a decision tree such as a binary tree to find a variable-value pair from the data trained. The process is recursive and repeating till the next maximum depth was reached or when the division of the subset could not be done anymore. Random forest works by following these metrics such as information gain , Gini impurity [3], and mean square [2] to estimate a new split. Random forest unites decision tree that is weak and turns it into a strong learner [3] A different method for predictions for aggregations can be used depending on the task [5].

3.4 Emotion classification based on a Hybrid of CNN and LSTM:

Many approaches combine CNN and LSTM models to capture both short-range and long-range dependencies [9], and in many cases, these combinations have provided SOTA systems. Chen et al. [36] improved accuracy in sentence-level sentiment classification by proposing a model containing a bidirectional LSTM and conditional random fields (BiLSTM-CRF) to classify the sentence type. The model also includes a 1- dimensional CNN to classify the emotion, tailored to specific sentence types by identifying the number of emotions expressed in each of the sentences. By dividing these sentences into groups, it reduces the complexity for each category which means, the learning algorithm has less complexity to deal with in the training data. This pipeline strategy obtains improved accuracy with the binary labeled version of Stanford sentiment treebank (SST) (88.3%) compared to 85.4% for Customer Reviews (CR).

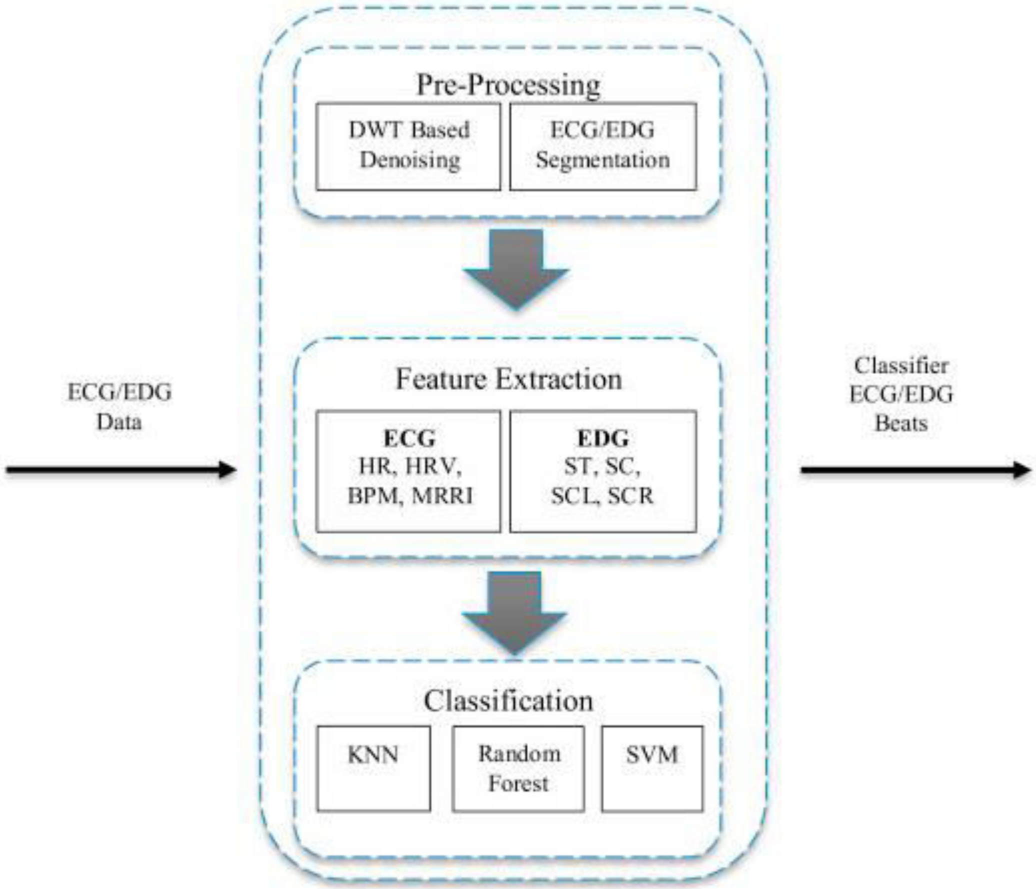

Fig 3:FE based classification architecture

4. Results Analysis

This section explores about results of various algorithms on different parameters to judge and defend existing works

Table 1:Results of various algorithms compared to proposed algorithm

Algorithm	Precission	Recall	F1 score	Accuracy (%)
SVM	71.52	72.35	73.58	81
RF	69.25	70.5	72.46	79
CNN	70.62	68.22	72.39	81
LSTM	73.77	59.37	73.95	83
CNN and LSTM	74.2	78.63	75.01	86

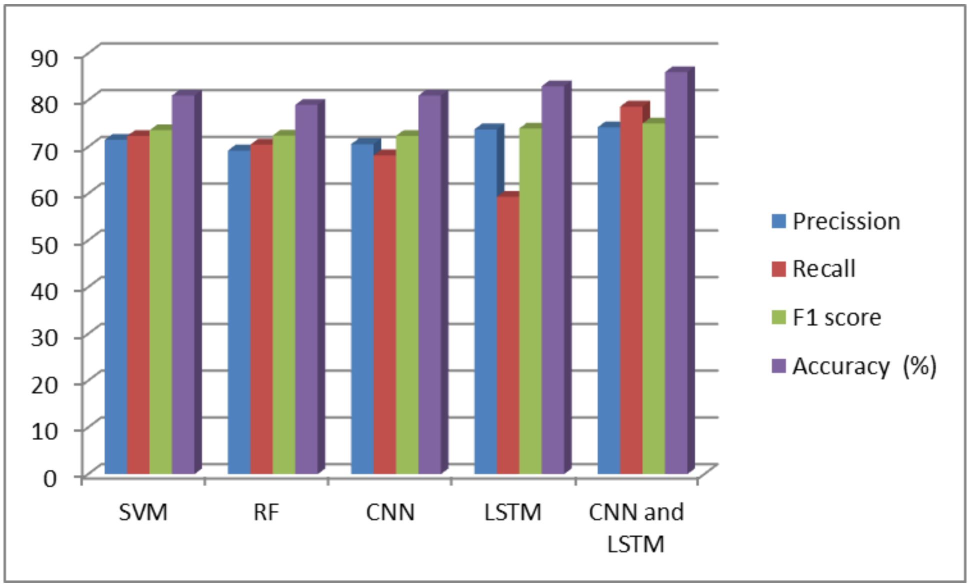

Fig 4: chart comparison of all parameters

Table 2: Algorithm processing time and detection rate

Algorithm	Processing Time(ms)	Detection Rate (%)
SVM	9.36	92.55
RF	10.12	91.65
CNN	9.57	93.47
LSTM	9.01	94.21
CNN and LSTM	8.95	95.33

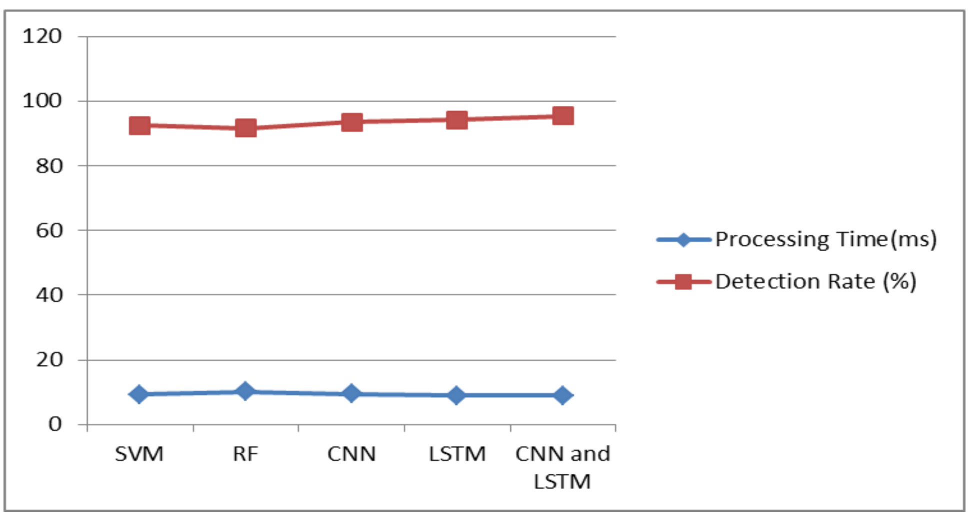

Fig 5: comparison chart for Algorithm processing time and detection rate

5. Conclusion

In this paper, we present a model for emotion recognition based on text analysis. Our proposed model integrates deep learning and machine learning methodologies, offering several advantages such as the ability to handle multi-text sentences, tweets, dialogues, keywords, and vocabulary while accurately identifying emotions. Additionally, we explore the characteristics of ECG and EDG signals and detail the equipment utilized for data collection, preprocessing, feature extraction, and signal classification. A crucial aspect of our discussion involves the classifiers employed and their effectiveness in analyzing different physiological signals. The classifiers utilized for emotion classification encompass SVM, cNN, and random forest, among others. Notably, SVM consistently yields the most reliable results, achieving an accuracy rate of 81% according to the machine learning classifier. On the other hand, in the deep learning realm, the LSTM model achieves the highest accuracy at 83%, while the CNN model attains the highest F1 score of 72.39%. Our hybrid model delivers a precision rate of 74.2%, recall rate of 78.63%, an F1 score of 75.01 and an overall precision rate of 74.2%.

References

1. A. A. Alnuaim, M. Zakariah, P. K. Shukla et al., "Human-computer interaction for recognizing speech emotions using multilayer perceptron classifier," *Journal of Healthcare Engineering*, vol. 2022, Article ID 6005446, 12 pages, 2022.
2. Nagesh, C., Chaganti, K.R. , Chaganti, S. , Khaleelullah, S., Naresh, P. and Hussan, M. 2023. Leveraging Machine Learning based Ensemble Time Series Prediction Model for Rainfall Using SVM, KNN and Advanced ARIMA+ E-GARCH. *International Journal on Recent and Innovation Trends in Computing and Communication*. 11, 7s (Jul. 2023), 353–358. DOI:https://doi.org/10.17762/ijritcc.v11i7s.7010.
3. S. Khaleelullah, P. Marry, P. Naresh, P. Srilatha, G. Sirisha and C. Nagesh, "A Framework for Design and Development of Message sharing using Open-Source Software," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 639-646, doi: 10.1109/ICSCDS56580.2023.10104679.
4. S. M. Mohammad and F. Bravo-Marquez, "WASSA-2017 Shared Task on Emotion Intensity," 2017, http://arxiv.org/abs/1708.03700
5. P. Xu, A. Madotto, C. S. Wu, J. H. Park, and P. Fung, "Emo2vec: learning generalized emotion representation by multi-task training," 2018, http://arxiv.org/abs/1809.04505.
6. W. Ragheb, J. Azé, S. Bringay, and M. Servajean, "Attention-based modeling for emotion detection and classification in textual conversations," 2019, http://arxiv.org/abs/1906.07020.
7. M. I. Thariq Hussan, D. Saidulu, P. T. Anitha, A. Manikandan and P. Naresh (2022), Object Detection and Recognition in Real Time Using Deep Learning for Visually Impaired People. IJEER 10(2), 80-86. DOI: 10.37391/IJEER.100205.
8. Suguna Ramadass and Shyamala Devi 2019 Prediction of Customer Attrition using Feature Extraction Techniques and its Performance Assessment through dissimilar Classifiers Springer's book series Learning and Analytics in Intelligent Systems, Springer.
9. Y. Zhang, Z. Zhang, D. Miao, and J. Wang, "Three-way enhanced convolutional neural networks for sentence-level sentiment classification," Information Sciences, vol. 477, pp. 55-64, 2019, doi: 10.1016/j.ins.2018.10.030.
10. P. Steno, A. Alsadoon, P. W. C. Prasad, T. Al-Dala'in, and O. H. Alsadoon, "A novel enhanced region proposal network and modified loss function: threat object detection in secure screening using deep learning," The Journal of Supercomputing, 2020/09/07 2020, doi: 10.1007/s11227- 020-03418-4.
11. J. Vinay Kumar, K. Shishir, and M. Prabhat, "Sentiment Recognition in Customer Reviews Using Deep Learning," International Journal of Enterprise Information Systems (IJEIS), vol. 14, no. 2, pp. 77-86, 2018, doi: 10.4018/IJEIS.2018040105.

12. C. Dhaoui, M. Webster Cynthia, and P. Tan Lay, "Social media sentiment analysis: lexicon versus machine learning," Journal of Consumer Marketing, vol. 34, no. 6, pp. 480-488, 2017, doi: 10.1108/JCM-03-2017-2141.
13. V.Krishna, Dr.V.P.C.Rao, P.Naresh, P.Rajyalakshmi " Incorporation of DCT and MSVQ to Enhance Image Compression Ratio of an image" International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Volume: 03 Issue: 03 | Mar-2016
14. P.Naresh,et.al., "Implementation of Map Reduce Based Clustering for Large Database in Cloud", Journal For Innovative Development in Pharmaceutical and Technical Science,vol.1,pp 1-4,2018.
15. D. Stojanovski, G. Strezoski, G. Madjarov, I. Dimitrovski, and I. Chorbev, "Deep neural network architecture for sentiment analysis and emotion identification of Twitter messages," Multimedia Tools and Applications, vol. 77, no. 24, pp. 32213-32242, 2018, doi: 10.1007/s11042-018-6168- 1. [15] L. Mou et al., "How Transferable are Neural Networks in NLP Applications?," arXiv.org, 2016.
16. S. Xiong, H. Lv, W. Zhao, and D. Ji, "Towards Twitter sentiment classification by multi-level sentiment-enriched word embeddings," Neurocomputing, vol. 275, pp. 2459-2466, 2018, doi: 10.1016/j.neucom.2017.11.023.
17. X. Sun, C. Li, and F. Ren, "Sentiment analysis for Chinese microblog based on deep neural networks with convolutional extension features," Neurocomputing, vol. 210, pp. 227-236, 2016, doi: 10.1016/j.neucom.2016.02.077.
18. M. Hanafy, M. I. Khalil, and H. M. Abbas, "Combining Classical and Deep Learning Methods for Twitter Sentiment Analysis," in Artificial Neural Networks in Pattern Recognition, Cham, L. Pancioni, F. Schwenker, and E. Trentin, Eds., 2018// 2018: Springer International Publishing, pp. 281-292.
19. X. Sun and C. Zhang, "Detecting anomalous emotion through big data from social networks based on a deep learning method," Multimedia Tools and Applications, vol. 79, no. 13-14, pp. 9687-9687, 2020, doi: 10.1007/s11042-018-5665-6.
20. B.M.G. Prasad, P. Naresh, V. Veeresh, "Frequent Temporal Patterns Mining With Relative Intervals", International Refereed Journal of Engineering and Science ,Volume 4, Issue 6 (June 2015), PP.153-156.
21. T. Aruna, P. Naresh, A. Rajeshwari, M. I. T. Hussan and K. G. Guptha, "Visualization and Prediction of Rainfall Using Deep Learning and Machine Learning Techniques," 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2022, pp. 910-914, doi: 10.1109/ICTACS56270.2022.9988553.
22. N. Azam, B. Tahir, and M. A. Mehmood, "Sentiment and emotion analysis of text: a survey on approaches and resources," in *Proceedings of the 7th International Conference on Language and Technology (CLT)*, pp. 87–94, IEEE, Lahore, Pakistan, February 2020
23. Hussan, M.I. & Reddy, G. & Anitha, P. & Kanagaraj, A. & Pannangi, Naresh. (2023). DDoS attack detection in IoT environment using optimized Elman recurrent neural networks based on chaotic bacterial colony optimization. Cluster Computing. 1-22. 10.1007/s10586-023-04187-4.
24. E. M. Onyema, P. K. Shukla, S. Dalal, M. N. Mathur, M. Zakariah, and B. Tiwari, "Enhancement of patient facial recognition through deep learning algorithm: ConvNet," *Journal of Healthcare Engineering*, vol. 2021, Article ID 5196000, 8 pages, 2021.
25. T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, http://arxiv.org/abs/1301.3781.
26. P. Naresh, P. Srinath, K. Akshit, M. S. S. Raju and P. VenkataTeja, "Decoding Network Anomalies using Supervised Machine Learning and Deep Learning Approaches," 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2023, pp. 1598-1603, doi: 10.1109/ICACRS58579.2023.10404866.
27. D. Ghazi, D. Inkpen, and S. Szpakowicz, "Detecting emotion stimuli in emotion-bearing sentences," *Computational Linguistics and Intelligent Text Processing*, Springer, Cham, pp. 152–165, 2015.
28. V. Krishna, Y. D. Solomon Raju, C. V. Raghavendran, P. Naresh and A. Rajesh, "Identification of Nutritional Deficiencies in Crops Using Machine Learning and Image Processing Techniques," 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2022, pp. 925-929, doi: 10.1109/ICIEM54221.2022.9853072.

29. B. Narsimha, Ch V Raghavendran, Pannangi Rajyalakshmi, G Kasi Reddy, M. Bhargavi and P. Naresh (2022), Cyber Defense in the Age of Artificial Intelligence and Machine Learning for Financial Fraud Detection Application. IJEER 10(2), 87-92. DOI: 10.37391/IJEER.100206
30. A. A. Alnuaim, M. Zakariah, C. Shashidhar et al., "Speaker gender recognition based on deep neural networks and ResNet50," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 4444388, 13 pages, 2022.