Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Emotion Recognition with Multi Physiological Signals: A Deep Learning Approach

Naveen Palanichamy[1,*], Su Cheng Haw[1], Revathi K[2], Kok-Why Ng[1], Suthent Tamilselvam[1]

[1] Faculty of Computing and Informatics, Multimedia University, Persiaran Multimedia, 63100, Malaysia.
[2] Department of Artificial Intelligence and Data Science, Dhanalakshmi College of Engineering, Chennai 601301, India

ARTICLE INFO

Keywords:

Emotion recognition; physiological signals; feature optimization; deep learning; machine learning

ABSTRACT

Human emotions, a complex interplay of psychological and physiological signals are a critical aspect of human interaction and well-being. Emotion recognition models in general capture human behaviour via facial features, voice/speech, and physiological signals, and evaluate and predict emotional states. The physiological signals, like brainwave, heart rate, eye movement, or galvanic skin response are the major cause of emotional changes. The combination of these signals contributes to the emotion change. Thus, effective models with combined physiological signals will serve better solutions compared to other modalities in practice. The rapid and precise recognition of emotions remains a challenge due to not enough combined physiological datasets. The promising characteristics of deep learning will help in achieving efficient emotional recognition models in place which use physiological signals as the predominant concern. The paper aims to propose a framework that utilizes Convolutional Neural Networks (CNN) with feature engineering to process combined physiological signals enabling the model to discern relevant emotional patterns while filtering out noise. To improve the consistency and accuracy of the prediction of emotions Linear Regression (LR) is used and Random Forest (RF) was employed to reduce the overfitting and noise problem. The efficacy of the proposed model was rigorously evaluated using standard performance metrics which are precision, recall, F1 score and accuracy. The proposed model, CNN+LR+RF with feature optimization model performed well with 61% accuracy compared to other proposed models.

1. Introduction

Human feelings are complex and highly individual; thus, they become the foundation of our behaviour and mental health. These emotions are all largely recognized, the physiological responses, cognitive evaluation, and cultural influences being the main influencers [1]. The emotions are joy, sadness, anger, fear, surprise, and disgust. A subjective method is based on personal opinions and feelings rather than objective measurements. Subjective methods were employed in the research to elicit emotions from participants and collected data for emotional analysis, selecting four individuals

* *Corresponding author.*
E-mail address: author.mail@gmail.com

https://doi.org/10.37934/araset.63.1.188205

aged 20-30 to monitor Photoplethysmography (PPG) signals while viewing music videos chosen for their valence and arousal levels based on a survey of 10 people. Even though they are subjective, the study of emotions has much to offer in human-computer interactions; hence, there is a need to find ways of substituting subjective methods with objective ones, such as combined physiological signals for emotional analysis [2]. Consequently, several physical signals, including the heart rate, skin conductance, facial expressions, and EEG (electroencephalogram), are used because they supplement one another and provide more information regarding a person's emotional state. The sophisticated model systems composed of such diverse signals are more likely to be precise in emotion recognition than the single-signal systems. This highlights the need for adaptive models that can handle minor variations in physiological responses [3].

Deep learning (DL) algorithms have proved to be most useful for recognizing emotions from physiological signals, resulting in more precise, robust and interpretable structures. The development of CNNs in pattern recognition has proved to be a significant milestone in DL and is currently being used for feature recognition in complex signals. Nevertheless, most of the studies that were conducted restricted physiological signals to only one or two signals. It can be challenging for emotion recognition models to overfit and reduce generalizability. As an example, authors took the opportunity in pupillometry to capture human emotion, disgust, and fear, which is directly and automatically processed by the brain [3]. Similarly, another study focused on the emotions of disillusion and fear [4]. In this study, to make an emotion state assertion, authors used heartbeat dynamics for assertion using a support vector machine (SVM) [5]. This research used 1D CNN for the categorization of short-term emotions based on single pulse photoplethysmogram (PPG) signals [2]. This study seeks to forecast arousal and valence by analyzing individual peripheral signals (galvanic skin response (GSR), RSP, BVP, Temp) through CNN [3]. The limitation in signal variety often results in CNN models underperforming, which has led researchers to integrate LR techniques to enhance prediction consistency and accuracy. Hence, it is necessary to have a model that can recognize minor variations in the inputs and predict complex emotions with reasonable accuracy.

Nevertheless, the potential of DL in emotion recognition remains substantial. The capability of combining and analyzing the signals from different physiological systems would result in more precise models that can recognize and identify a wider variety of emotional states [5]. The primary objective of this research is to propose a framework based on CNN to improve efficiency and accuracy of emotion prediction on combined physiological signals. Additionally, this research seeks to address the challenge of overfitting and noise in emotion recognition models.

The structure of this paper is as follows: Section 1.1 lists relevant research done in recent years in emotion recognition using physiological signals. Section 2 explains the techniques used in this research. Section 3 discusses the results of this research, and Section 4 contains the conclusion of this research.

1.1 Literature Review

The literature review shows the research on face recognition and the employment of machine learning (ML) and DL algorithms for emotion recognition. It includes research areas that introduce an application of modern techniques for increasing precision in natural language processing; for example, feature optimization methods and DL models with CNNs and SVMs for emotion sensing are used from EEG signals. This evidences the ability of DL to recognize fine nuances of emotions from physiological data, where emotions are usually represented in a complex way.

1.1.1 Feature optimization

The paper introduced a new approach to feature optimization [6]. Targeted at its compatibility with classification techniques that seek to reduce noise and enhance the quality of data in natural language processing, it outperformed basic models. At the same time, standard Particle Swarm Optimization (PSO) methods might be mute about particular elements necessary for the process of emotion recognition and therefore offer an inferior outcome in comparison to the proposed time adaptive PSO version, which suggests that conventional PSO algorithms might be incapable of capturing all necessary features for accurate emotion recognition. In this article, the authors directed their attention to two issues the forest optimization method has: control parameter setting and population generation, which are primarily addressed and lead to improved classification accuracy and dimension reduction respectively [7]. The effectiveness of PSO in searching the solution space stems from the power of the swarm to work as a coherent unit in order to achieve a goal, but it also comes with the problem of high dimensional feature space, which entails high computations. A tweaked version of PSO has been utilized to determine the feelings in the frames of videos to include a time constant in the algorithm to curtail the premature convergence and enable the identification of emotional expressions across various datasets such as Cohn-Kanade (CK+), JAFFE, and NIMH-ChEFS [8].

An innovative two-stage feature optimization procedure, which includes the integration of the FCA and ReliefF algorithm, has significantly boosted emotion recognition from physiological signals using the DEAP database and the Augsburg Bio-Signal Toolbox by refining the feature set through discarding of unimportant features [9]. Likewise, a dual combination is implemented whereby recursive feature elimination (RFE) is used in conjunction with several feature importance methods – support vector machine (SVM), RFand generalized boosted regression models for data with large numbers of features [10]. This operation eliminates noise and overfitting, which in turn increases model accuracy. This method underscores the need for solutions to be scalable across various application scenarios in those cases when such a process could be tailored to specific application needs to be most effective. In contrast, some methods may be specific to a particular dataset.

In addition, ensemble classifiers allow RFE to be integrated into their design to increase stroke prediction accuracy and reduce overfitting by using several base classifiers simultaneously [11]. To increase generalization and decrease noise in the model, IGRF-RFE [12-14] is a method that uses a combination of information gain, RF importance, and RFE for feature selection in network intrusion detection. On the other hand, these studies advance the issues of feature enhancement and emotion detection, as well as the problems that still need to be dealt with. Most of these techniques need to be modified explicitly for various requirements, which is unrealistic in non-mass application areas. Moreover, the accuracy of these models, which depend on vast and fortuitous datasets, presents issues in their generalizability and applicability to different real-world scenarios.

The comparative analysis of PSO with RFE is presented in Table 1. This article focuses on Feature optimization algorithms for emotion recognition and other related tasks. The results of the new feature optimization approach used in NLP tasks appear promising due to noise elimination. At the same time, traditional PSO algorithms tend to miss crucial parts regarding emotion classification. An improved variant of PSO with a time constant converged on the problem of premature convergence in multiple emotion recognition or expression identification tasks using CK+, JAFFE and NIMH-ChEFS datasets. It achieved better performance metrics on forest optimization algorithms' classification accuracy and dimension reduction. In conclusion, these results highlight not only the progress but also the problems of feature optimization, pointing out the need for scalable and adaptable approaches for multiple datasets.

Table 1
Comparison of Particle Swarm Optimization (PSO) and Recursive Feature Elimination (RFE) usage in various studies

Study	Particle Swarm Optimization (PSO)	Recursive Feature Elimination (RFE)
[4]	✓	
[5]	✓	
[6]	✓	
[7]	✓	
[8]		✓
[9]		✓
[10]		✓
[11]		✓
[12]		✓

1.1.2 Machine learning challenges

ML is crucially essential in emotion recognition, which is achieved with the help of the identification of complex human emotional expressions using physiological and behavioural cues. Recent improvements in DL have given rise to novel models that exploit CNN and support vector machines (SVM) for emotion recognition from EEG signals [15]. This method has proven its potential by translating EEG signals into 2D spectrograms to increase recognition accuracy for diverse datasets such as DEAP and SEED. Additionally, the performance of the subject-independent classification strategies can be optimized, as it was noted that the individual baselines exhibited some variability.

The study further elaborates on the effectiveness of CNNs in EEG-based recognition of emotions compared to RF algorithms [16]. The research indicated that CNN obtained an accuracy of 67%, which signals their proficiency in managing intricate data interpretations and implies optimizations could span a broader category of emotions. In addition, class EEG physiologic signals integration with classes has been explored, and such multimodal data fusion has been found to significantly boost the accuracy of emotion recognition of the DEAP platform [17]. Further advances are needed in developing particularly adaptive algorithms required to process data from diverse sources.

A multi-featured deep forest model has been analyzed deeply in the previous research [18]. The model utilized the PSD and DE from the DEAP dataset and achieved an average accuracy of 71%. The model performed better than the traditional classifiers such as RF, SVM and KNN. This model marked the first application of DL. However, the authors of the paper tried to reproduce the results of Cheng *et al.,* [19], but this time using different methods to induce emotions and confirmed that FFNN had 70% accuracy. The effectiveness of DL technology has been confirmed. Research using RF obtained the lowest accuracy of 62% [20]. The DEAP dataset includes over 50 videos labelled with affective dimensions that range from happy to sad and calm to angry, while the other 58% of videos exhibit emotions that are less clearly definable. It can be concluded that algorithmic approaches to emotion recognition still need to be developed to achieve higher accuracy and generalization. The study investigates how emotions can be identified from Hindi audio using different ML methods with significant accuracy using the RF algorithm [21].

Though these studies indicate some encouraging developments, certain constraints can circumscribe the more extensive application of these machine-learning models to recognize emotions. Additionally, the diversity in the accuracy of different subjects with the classifiers suggests the need for better and more objective approaches to classification. Furthermore, although fusing multiple forms of data seems promising, existing models need to be improved for better handling of the variety and complexity of data types to achieve uniform and reproducible outcomes across various datasets.

1.1.3. Deep learning approaches

DL techniques have significantly impacted emotion detection, especially in the case of modern neural systems. Emotion recognition from EEG signals was done with a randomized CNN model, which tagged a greater accuracy than existing systems on the DEAP dataset [4]. Similarly, as described in Sharma *et al.,* [21], negative feelings are accurately identified by an LSTM architecture from EEG data in the DEAP and SEED datasets. DL can analyze and extract complicated emotional patterns from physiological data.

Several ML approaches, including SVM, LR, CNN, and Recurrent Neural Networks (RNN), have been thoroughly investigated in the context of the recognition of emotions from EEG signals, for example, in Nakisa *et al.,* [23]. The research further identifies the challenges faced in emotion recognition from data. It stresses the need to select suitable features and classifiers to increase the precision. In addition, CNN has been shown to successfully classify emotions by analyzing signals from electrodermal activity in the MAHNOB and DEAP datasets, with 78% and 82% classification accuracy, respectively. This research fails to demonstrate CNN's problem-solving capabilities. It also highlights the various issues of building models that could yield good results for different individuals and datasets.

A recent approach generally integrates early and late fusion models into the time-based multimodal learning model using multi-signal interfaces [24]. The time-based model is illustrated particularly with the multimodal model and, in this instance, BVP and EEG sensors acquired from devices. This suggests that the fusion technique has great potential in overcoming the weakness of short-tersignalls operations by improving the accuracy of the emotion recognition systems.

In addition, the investigation on expanding CNN to identify six primary emotions from physiological signals has pointed out the necessity to emphasize data preprocessing and tuning of parameters using adequate techniques for best results [29]. The reports focused on cross-subject emotion recognition where pre-trained CNNs are applied, which implies the potentiality of more generalization, but the data limits the work [30]. The research on different architectures of DL approaches for analyzing emotions and stress indicates the need to validate and optimize these models in hyperparameter comparison studies [31]. Besides, the BI-LSTM model application on EEG signals for emotion recognition has attained appreciable success [32]. However, the employed computational resources and the complexity of tuning such models preclude their use in real-time situations. The current study has developed a stress recognition model that employs CNNs to extract features from video and EEG data and uses XGBoost for classification [33]. It has been demonstrated that this model can accurately detect stress and work better than conventional methods like decision trees (DT) and RF. The systematic review highlights the application of RFE and CNNs in the task of Alzheimer's disease diagnosis [34].

Table 2 shows the DL approaches in recent studies. Although DL for emotion recognition has progressed greatly, some limitations still exist. One of the disadvantages of training complex models, including CNN and extended short-term networks (LSTM), is the necessity of massive, well-annotated datasets, especially in situations where data privacy is significant and the data collection process is complicated. Besides, most of these models are computationally expensive and complex; thus, implementation in real-time systems is very challenging. As such, the issue is finding ways to develop these models to make them more efficient and flexible.

Table 2
Deep Learning Approaches used in recent studies

Study	Models		
	CNN	LSTM	RNN
[19]	✓		
[20]		✓	
[21]	✓		✓
[22]	✓		
[23]	✓		
[24]		✓	
[25]	✓		
[26]	✓		
[27]	✓		
[28]	✓		
[29]	✓		
[30]	✓		
[31]		✓	
[32]		✓	
[33]	✓		
[34]	✓	✓	

2. Methodology

This section of the paper discusses the methodology of emotion recognition using DL and physiological signals to provide a broad description of the methods and technologies used. The purpose is to develop a technique that can detect and process a wide variety of physiological signals covering human emotions, using innovative machine-learning methods to achieve high accuracy in detecting emotions. Figure 1 shows the proposed methodology.

2.1 Data Obtained

The paper uses the ASCERTAIN dataset, a group of data collected from 58 participants who watched 36 emotional videos and rated their feelings on a seven-point scale ranging from very boring to very exciting [35]. The bio-signals of the participants were obtained through commercial sensors that were used to record physiological signals. The Electroencephalogram (EEG) captures brain activity, the Electrocardiogram (ECG) monitors heart rate, galvanic skin response (GSR) to track sweat gland response and the EMO to track facial response. A total of 36 emotional videos were presented. Table 3 shows the dataset description.

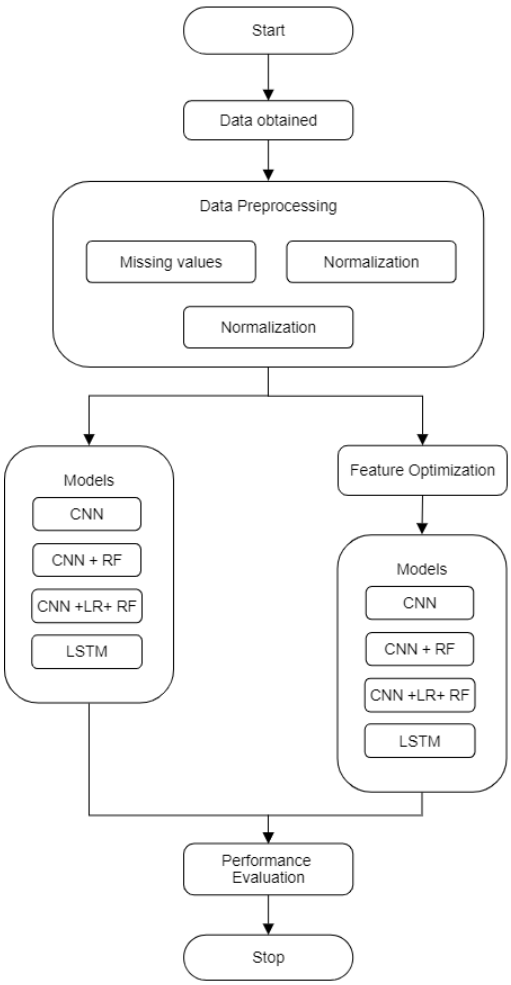

Fig. 1. Proposed Methodology

Table 3
Dataset Description

Features	Description
Participants	Number of participants
Physiological Signals	ECG recordings in combination with other physiological activities.
Video Stimuli	Classified video clips according to emotional valence and arousal levels.
Emotional Categories	Subcategories under the umbrella of emotional categories
Personality Traits	Big Five personality characteristics
Data Collection Method	Physiological data collected from the participants during the video stimuli sessions.

2.2 Data Pre-Processing

The ASCERTAIN dataset undergoes rigorous preprocessing steps to conform the to the input data format specified by the CNN. It starts by merging features from different providers of physiological data into a single feature matrix known as 'X'. To deal with the differences in sensor data outputs, we implement normalization processes with the use of 'StandardScaler' so that every feature has a mean of zero and a standard deviation of one. This step is significant in improving the model's convergence and in reducing the effects of sensor variations.

Also, the occurrence of infinite numbers in 'X' is replaced with NaNs that are later filled with the mean value to prevent data loss. By validating and normalizing the data set, we improve the performance reliability and accuracy of our model on different devices. During the development phase, we used a variety of libraries and frameworks in constructing our model to cope with different types and configurations of sensors. This strategy stems the variations from the sensor data and also gives a robust implementation for emotion identification based on biological signals using DL methodologies.

2.3 Feature Optimization

RFE and PSO are the feature optimization techniques considered to fine-tune the features. RFE runs through the existing subset and eliminates the least important features, allowing the model to be built on minimal features and finding the optimum number of features. It assists in increasing the quality of the models by cleaning the data and decreasing the model complexity, reducing the chances of overfitting. In the same way, PSO can efficiently search the feature space and choose relevant feature subsets depending on principles of swarm intelligence and improve the model action. This iterative fine-tuning weighs more for the model to learn better generalization from the training dataset towards predicting emotional emotions from physical signals and emotional state recognition patterns.

2.4 Model Building

The use of a CNN in this study is because of its ability to recognize spatial features in physiological signals, which is important in emotion recognition tasks. The model is trained based on the dataset, of which the training and validation data ratio is 80:20 to allow robustness and generalization. The activation of the second to the last layer will be the feature representations of the physiological signals to which the training features and the testing data will be extracted using the CNN model, which will extract the implicit, hidden patterns in the multimodal data, resulting in accurate emotion recognition. Figure 2 shows the CNN architecture.

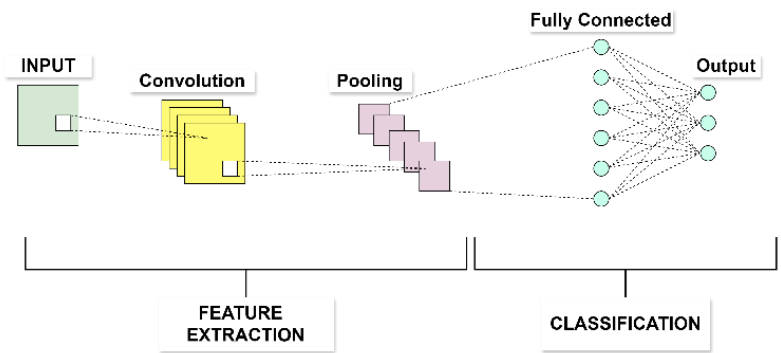

Fig. 2. Visualize of CNN architecture

2.5 Evaluation Metrics

To evaluate the performance evaluation, certain evaluation metrics will be used, including precision, accuracy, F1 score, recall, and Receiver Operating Characteristic (ROC).

2.5.1 Precision

Precision is a fundamental metric that calculates the percentage of true positives among all anticipated positives. High accuracy leads to confidence in the models' favourable forecasts, but low precision leads to the need for a change. The precision is computed as follows:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{1}$$

2.5.2 Accuracy

Accuracy in performance evaluation is defined as the proportion of correctly classified instances among all occurrences in the dataset. It is a common metric for evaluating the overall performance of a classification model. Accuracy is computed as follows:

$$Accuracy = \frac{Number\ of\ correctly\ predicted\ emotions}{Total\ number\ of\ predictions} \tag{2}$$

2.5.3 Recall

Recall, often known as the true positive rate, is a performance metric used in classification tasks to assess how successfully a model recognises positive cases. It represents the proportion of accurately anticipated positive events to all actual positive instances. Recall is computed as follows:

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{3}$$

2.5.4 F1 score

The F1 score is a metric for evaluating the performance of a classification model. The harmonic mean of accuracy and recall generates a single score that balances both measurements. The F1 score is computed as follows:

$$F1\ score = \frac{2*(Precision + Recall)}{Precision + Recall} \tag{4}$$

2.5.5 Receiver Operating Characteristics (ROC)

The ROC graph shows false positive rates on the X axis and true positive rates on the Y axis. The False Positive Rate and the True Positive Rate are computed as follows:

$$True\ Positive\ Rate = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{5}$$

$$False\ Positive\ Rate = \frac{False\ Positive}{False\ Positve + True\ Negative} \tag{6}$$

3. Results And Discussion

The proposed model was trained and tested with ASCERTAIN dataset. In the beginning, we used CNN both with and without feature optimization to get the accuracy of the model. The models' performance was not good because of the dataset with both linear and nonlinear relationships

between the variables. Nevertheless, the feature optimization model of CNN outperforms the one without feature optimization.

3.1 Convolutional Neural Network (CNN) and Feature optimization

Initially, CNN architecture was designed with two 1D convolutional layers with ReLU activation functions, followed by max pooling layers to downsample the feature maps. Then, a flattened layer is added to transform the 3D output into a 1D vector, followed by two Dense layers with ReLU and softmax activation functions to output probabilities for each emotion class. Next, the model is instantiated and trained using the training data. An early stopping callback is incorporated to monitor the validation loss and restore the best weights when the loss fails to improve for 10 epochs.

In this study, we assessed the performance of two CNN models to classify the emotions into seven classes. The models implemented were CNN and CNN with feature optimization, including RFE and PSO. We selected RFE as one of the feature selection algorithms because of its effectiveness in boosting model performance by identifying and removing fewer essential features from the model and targeting more relevant ones. PSO was also examined as an alternative method due to its capacity to effectively search the feature space and select feature subsets with the help of swarm intelligence principles, which could increase model accuracy. The CNN+RFE model resulted in an accuracy of 28%, a notable increase over the regular CNN model's performance, while PSO achieved 19% accuracy.

Based on the given Table 3, the optimized model showed balanced improvement in all classes, especially in class 4, where the precision is 0.32 and recall is 0. 47. This indicates that the CNN model with the feature optimization gives better results, with an overall accuracy of 28% than the 18% of CNN model. It highlights the role of feature optimization in enhancing model accuracy and its efficiency for multi-class prediction. It also indicates that these enhancements are critical to improve the performance of CNN in classification tasks. As the RFE outperformed PSO, the feature optimization used with CNN is RFE for the other models discussed in this paper.

3.2 CNN + RF

Next, CNN with a RF model was used to improve the prediction accuracy by reducing overfitting, which is an issue of CNN. We assessed the performance of CNN+RF and CNN+RF with feature optimization. The results are shown in Table 4.

Both models show high precision levels for all classes, but there is a significant difference in recall rate, which considerably impacts the F1 scores. From the CNN+RF with feature optimized model, the accuracy increased slightly from 28% to 30%, and in most of the cases, it has better recall and F1-score, especially for the class 0 (with an F1-score of 0.91).

3.3 CNN + LR + RF

To improve the consistency and accuracy of the prediction, LR was included. The integrated model utilized CNN for feature extraction, followed by LR and RF models for prediction, which was trained and tested and achieved notable performance on the physiological dataset. It is because CNN is excellent at automatically learning relevant features from the raw data and can capture intricate patterns and relationships within the signals crucial for recognizing different emotions. Feature optimization technique was applied to identify the most informative subset of features. CNN excels

in feature extraction from datasets and learning spatial patterns, providing rich features for emotion complexity.

Table 3
Comparative analysis of CNN model metrics with and without feature optimization

Model	Accuracy	Class	Precision	Recall	F1-score
CNN	18%	0	0.00	0.00	0.00
		1	0.17	0.05	0.08
		2	0.00	0.00	0.00
		3	0.20	0.82	0.32
		4	0.50	0.02	0.03
		5	0.12	0.01	0.02
		6	0.11	0.10	0.10
CNN with RFE	28%	0	0.12	0.11	0.12
		1	0.12	0.10	0.11
		2	0.14	0.11	0.12
		3	0.26	0.25	0.25
		4	0.32	0.47	0.38
		5	0.31	0.27	0.29
		6	0.25	0.05	0.08
CNN with PSO	19%	0	0.00	0.00	0.00
		1	0.04	0.15	0.06
		2	0.09	0.21	0.13
		3	0.20	0.07	0.11
		4	0.35	0.18	0.24
		5	0.25	0.25	0.25
		6	0.29	0.05	0.08

Table 4
Comparative analysis of CNN-RF model metrics with and without optimization

Model	Accuracy	Class	Precision	Recall	F1-score
CNN+RF without feature optimization	28%	0	1.00	0.86	0.91
		1	1.00	0.40	0.57
		2	1.00	0.23	0.37
		3	1.00	0.05	0.09
		4	0.00	0.00	0.00
		5	1.00	0.01	0.03
		6	1.00	0.38	0.55
CNN+RF with feature optimization	30%	0	1.00	0.84	0.91
		1	1.00	0.45	0.62
		2	1.00	0.27	0.43
		3	1.00	0.05	0.09
		4	0.00	0.00	0.00
		5	1.00	0.01	0.01
		6	1.00	0.41	0.58

LR efficiently makes predictions based on these features, ensuring effective utilization. This integration brings several benefits, such as feature extraction and interpretable relationships leading to consistent and accurate predictions and computational efficiency for scalability and faster training

and prediction. The LR model was trained with and without optimized features to learn the coefficients that best fit the relationship between features and emotions.

For further enhancement of the model's accuracy, RF aggregates predictions from multiple DT, reduces overfitting and handles noise in the data. We applied feature optimization techniques to identify the most informative subset of features. It can reduce the dimensionality of the data by removing redundant and irrelevant features, which in turn improves the model's performance. RFE was used to select the top features, which ensured that the most informative features were used for training the LR and RF models. We assessed the performance of CNN+LR+RF and CNN+LR+RF with feature optimization. The performance results of the methods are shown in Table 5.

By comparing these two models, CNN+LR+RF with feature optimization gives better results than CNN+LR+RF without feature optimization. This implies that models without feature optimization lead to overfitting, which means the model will capture noises rather than underlying patterns, which can affect the model's performance.

Table 5
Comparative analysis of CNN-LR-RF model metrics with and without optimization

Model	Accuracy	Class	Precision	Recall	F1-score
CNN+LR+RF without feature optimization	52%	0	0.99	0.99	0.99
		1	0.93	0.82	0.88
		2	0.96	0.70	0.81
		3	0.88	0.30	0.44
		4	0.73	0.06	0.10
		5	0.88	0.14	0.25
		6	0.97	0.72	0.83
CNN+LR+RF with feature optimization	61%	0	0.97	0.99	0.98
		1	0.91	0.92	0.91
		2	0.90	0.81	0.85
		3	0.68	0.34	0.45
		4	0.53	0.17	0.25
		5	0.68	0.29	0.41
		6	0.91	0.82	0.86

3.4 LSTM

For a more comprehensive analysis of our proposed model's performance, we tested LSTM Networks, which serve as a benchmark for both their efficiency and the performance of the models we advance.

The efficacy of the Long Short-Term Memory (LSTM) model was tested under two conditions, without and with feature optimization through RFE. The results are shown in Table 6. In the case without feature optimization, the LSTM achieved an accuracy of 46%, showing commendable classification performance for Class 0 with a precision of 0.77, recall of 0.93 and F1-score of 0.84, which reliably identifies this class. The model, however, faced issues with the other classes, including Class 4, where a precision of 0.37 was achieved alongside its recall, which stood at only 0.21. However, applying feature optimization resulted in a PSP of the model accuracy of just 23%. This suggested that important features were possibly eliminated during the optimization process, an observation further confirmed by the decline in class metrics. For example, a fall in the Class 0 Precision to 0.30 coupled with the recall of 0.64, which provided an F1 score of 0.41. Such declines were witnessed in the performance of the rest of the classes as well, the most notable being Class 4, where precision and recall were around 0.04 and 0.01, respectively.

3.5 Discussion

The strengths and weaknesses of the model CNN+LR+RF in predicting multiclass outcomes will be discussed in this section.

We understand that it is essential to include a complete comparative evaluation to validate our choice of LR and RF models in conjunction with the CNN model. The addition of LR helps to interpret the obtained linear patterns from the features extracted by the CNN, thereby aiding in correlating the emotion of a subject to physiological changes. On the other hand, RF efficiently handles non-linear relations and, at the same time, minimizes overfitting, which is particularly essential in datasets with deep neural networks. This combination takes the best of both approaches to enhance prediction ability.

Table 6
Comparative analysis of LSTM model metrics with and without optimization

Model	Accuracy	Class	Precision	Recall	F1-score
LSTM without feature optimization	46%	0	0.77	0.93	0.84
		1	0.43	0.60	0.50
		2	0.43	0.43	0.43
		3	0.31	0.31	0.31
		4	0.37	0.21	0.27
		5	0.34	0.22	0.27
		6	0.44	0.56	0.49
LSTM with feature optimization	23%	0	0.30	0.64	0.41
		1	0.19	0.14	0.16
		2	0.22	0.34	0.27
		3	0.20	0.12	0.15
		4	0.04	0.01	0.01
		5	0.21	0.13	0.16
		6	0.22	0.24	0.23

Compared to the CNN + RF model, our CNN + LR + RF results, as reported in Table 5, are more trustworthy. More specifically, the CNN + LR + RF model accomplished a 52% accuracy score, surpassing the CNN + RF model by 24%. The integration of LR improves the interpretability of the features that CNN has extracted. It helps to understand the relationship between the emotional states and the physiological signals. Moreover, feature optimization is an effective means of improving the performance of the model as it assists in identifying and retaining critical features whilst reducing overfitting. Such optimization also enables advancement in accuracy to 61% of the CNN + LR + RF model. On the contrary, the CNN + RF models, both without and with feature optimization, struggle with accuracy, achieving only 28% and 30% accuracy, respectively. Moreover, with feature optimization, the CNN + LR + RF model reached 61% accuracy, which is considerably 31% larger than what CNN + RF + feature optimization could reach.

Figure 3 shows the training and validation accuracy of CNN+LR+RF with a feature optimization model. The training accuracy of this model increases, which indicates that the model is learning. The validation accuracy also increases, but it has no change after a certain point, which the model may not improve during training. However, it appears that CNN+LR+RF with a feature optimization model is learning and fitting to the training data.

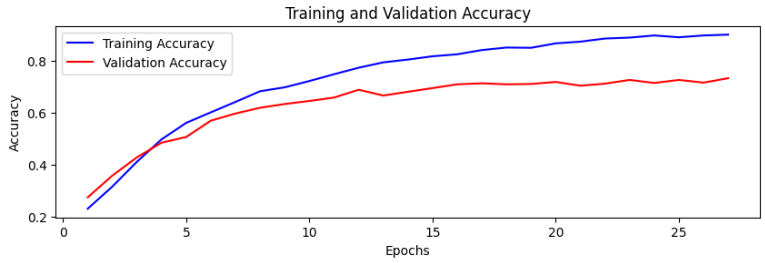

Fig. 3. Training and validation accuracy over Epochs

Moreover, our evaluation of ROC curves and AUC scores provides further support to these conclusions. CNN+RF, together with RFE, as shown in Figure 4, can achieve AUC scores between 0.48 (Class 0) and 0.57 (Class 6). Even though it can increase accuracy, the costs in computational resources do not offer any benefits to the overall process. Meanwhile, the combination of CNN+LR+RF, which combines feature optimization, achieves AUC scores close to ideal separation among multiple types with Class 0 (AUC = 1.00), Class 1 (AUC = 0.98), and Class 6 (AUC = 0.97). The AUC indicators of CNN+LR+RF with the RFE model, as shown in Figure 5, show greater potential in discriminating between various emotional conditions compared to CNN+RF with the RFE model. This indicates that our combined framework has an increased capacity to extract the features necessary for achieving accurate emotion recognition. In summary, these conclusions confirm our methodological approach and demonstrate the advantages of handling the emotional recognition problem by combining LR and RF with the CNN architecture.

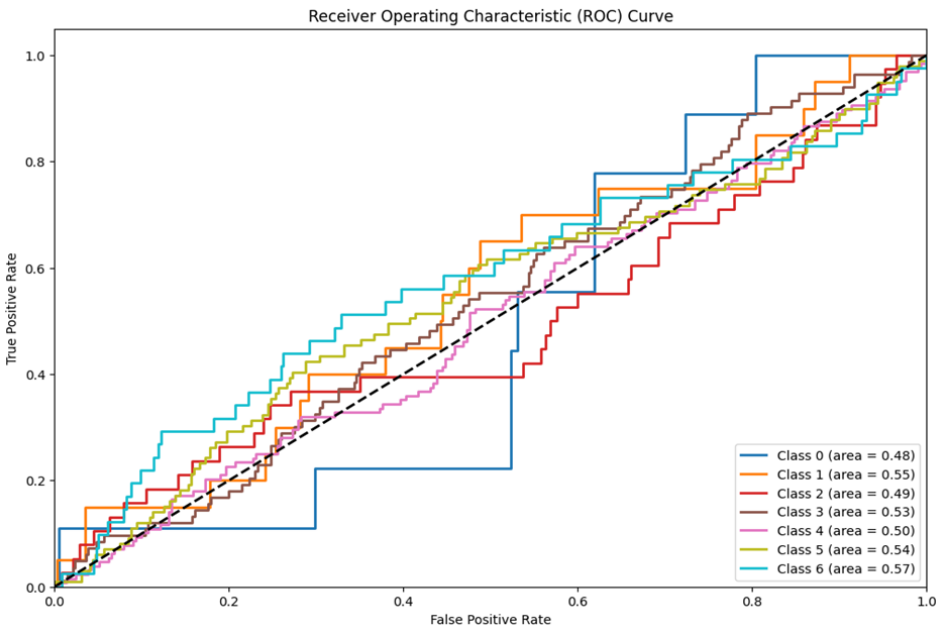

Fig. 4. CNN+RF+RFE model - Receiver Operating Characteristic (ROC) Curves for Multi-class Classification

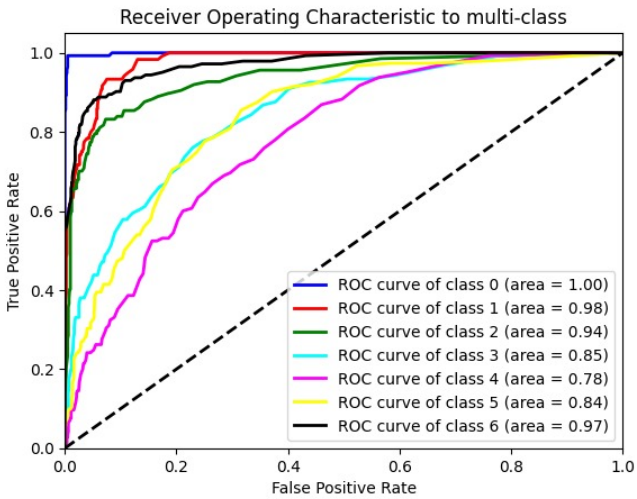

Fig. 5. CNN+LR+RF+RFE model - Receiver Operating Characteristic (ROC) curves for multi-class classification

The comprehensive emotion recognition models are depicted in Figure 6, with CNN achieving a fair baseline. Rationally, the proposed approach moves from classification tasks to regression tasks by augmenting RF to make it the CNN+RF model. In this model, better performance is obtained because it addresses non-linear relationships, thereby controlling overfitting. Then, adding LR to the model, CNN+LR+RF improves the robustness of the model while maintaining decent accuracy. The accuracy of CNN+LR+RF increases to 61% with the integration of RFE. The LSTM models suffer from poor results, translating into recognition accuracy of only 46% without feature improvement and 23% with, which further confirms the effectiveness of the CNN+LR+RF+RFE model for these tasks.

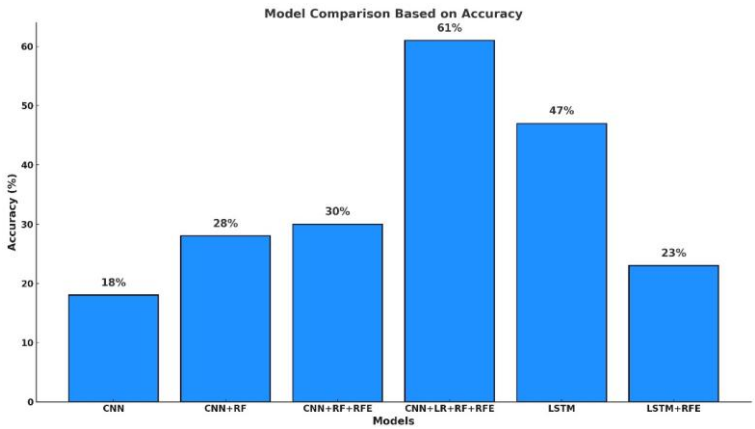

Fig. 6. Models comparison

Overall, CNN+LR+RF model results are more reliable compared to LSTM models. LR was integrated with CNN and RF because it can improve the consistency and accuracy of the predictions and leverage the power of CNNs to extract more meaningful features from the dataset. We obtained the accuracy of this model, which is 61%, 31% more compared to CNN+RF. Besides that, feature optimization also enhances the model's performance, identifying the most important features and reducing the overfitting issue.

4. Conclusion

Human emotions, a complex blend of psychological and physiological cues, play a vital role in human well-being and social interactions. Emotion recognition models use facial expressions, voice and physiological signals such as brainwaves and heart rate to anticipate emotional states, with the later playing a significant role in emotional fluctuations. Integrating these signals offers more effective solutions for emotion recognition models than other modalities. The study proposed a framework combining CNN and feature engineering to analyze physiological signals for emotional pattern recognition, utilizing LR for consistent emotion prediction and RF to address overfitting issues. According to the results of the study, the CNN+LR+RF with feature optimization has proved to be an effective model for the recognition of psychological states that can be determined by physiological signals. Including feature optimization approaches has become exceedingly helpful in increasing the accuracy of the models using the systematic tuning of the data input and eliminating overfitting and noise.

Acknowledgement
This research was funded by a grant from Multimedia University and Dhanalakshmi College of Engineering (International Matching Grant MMUE/230042).

References
[1] Oh, SeungJun, Jun-Young Lee, and Dong Keun Kim. "The design of CNN architectures for optimal six basic emotion classification using multiple physiological signals." *Sensors* 20, no. 3 (2020): 866. https://doi.org/10.3390/s20030866

[2] Lee, Min Seop, Yun Kyu Lee, Dong Sung Pae, Myo Taeg Lim, Dong Won Kim, and Tae Koo Kang. "Fast emotion recognition based on single pulse PPG signal with convolutional neural network." *Applied Sciences* 9, no. 16 (2019): 3355. https://doi.org/10.3390/app9163355

[3] M. Pidgeon, N. Kanwal, N. Murray, and M. Asghar, "End-To-End Emotion Recognition using Peripheral Physiological Signals," *in 35th British HCI Conference Towards a Human-Centred Digital Society*, HCI 2022, BCS Learning and Development Ltd., Jul. 2022. https://doi.org//10.14236/ewic/HCI2022.19

[4] Z. Deng, Z. Shi, Z. Wang, and T. Liu, "Research on Feature Optimization Scheme Based on Data Feature Enhancement," *in Proceedings - 2021 21st International Conference on Software Quality, Reliability and Security Companion*, QRS-C 2021, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 270–278. https://doi.org//10.1109/QRS-C55045.2021.00048

[5] Q. Xie, G. Cheng, X. Zhang, and P. Lei, "Feature selection using improved forest optimization algorithm," *Information Technology and Control*, vol. 49, no. 2, pp. 289–301, 2020, https://doi.org/10.5755/j01.itc.49.2.24858

[6] H. A. Shehu, W. Browne, and H. Eisenbarth, "Particle Swarm Optimization for Feature Selection in Emotion Categorization," *in 2021 IEEE Congress on Evolutionary Computation*, CEC 2021 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 752–759. https://doi.org//10.1109/CEC45853.2021.9504986

[7] L. Pan, S. Wang, Z. Yin, and A. Song, "Recognition of Human Inner Emotion Based on Two-Stage FCA-ReliefF Feature Optimization," *Information Technology and Control*, vol. 51, no. 1, pp. 32–47, Mar. 2022, https://doi.org/10.5755/j01.itc.51.1.29430

[8] H. Jeon and S. Oh, "Hybrid-recursive feature elimination for efficient feature selection," *Applied Sciences (Switzerland),* vol. 10, no. 9, May 2020, https://doi.org/10.3390/app10093211

[9] Mitra, Pooja, Sheshang Degadwala, and Dhairya Vyas. "Ensemble Classifier for Stroke Prediction with Recurshive Feature Elimination." (2023).https://doi.org//10.32628/CSEIT2390430

[10] Yin, Yuhua, Julian Jang-Jaccard, Wen Xu, Amardeep Singh, Jinting Zhu, Fariza Sabrina, and Jin Kwak. "IGRF-RFE: a hybrid feature selection method for MLP-based network intrusion detection on UNSW-NB15 dataset." *Journal of Big data* 10, no. 1 (2023): 15.https://doi.org//10.1186/s40537-023-00694-8.

[11] J. W. Kasubi and M. D. Huchaiah, "A Comparative Study of Feature Selection Methods for Activity Recognition in the Smart Home Environment," *in International Conference on Automation*, Computing and Renewable Systems, ICACRS

2022 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 890–895. https://doi.org//10.1109/ICACRS55517.2022.10029133

[12] Shan, Yongfu, Chen Xi, and Hui Wang. "Optimal Recursive Reature Selection Method Based on Random Forest." In *2023 IEEE International Conference on Image Processing and Computer Applications (ICIPCA)*, pp. 268-273. IEEE, 2023.https://doi.org//10.1109/ICIPCA59209.2023.10257890

[13] Pusarla, Nalini, Anurag Singh, and Shrivishal Tripathi. "Learning DenseNet features from EEG based spectrograms for subject independent emotion recognition." *Biomedical signal processing and control* 74 (2022): 103485.https://doi.org//10.1016/j.bspc.2022.103485

[14] V. B. Savinov, S. A. Botman, V. V. Sapunov, V. A. Petrov, I. G. Samusev, and N. N. Shusharina, "Electroencephalogram-based emotion recognition using a convolutional neural network," *Bulletin of Russian State Medical University*, vol. 8, no. 3, pp. 32–35, 2019, https://doi.org//10.24075/brsmu.2019.037

[15] Y. Fang, H. Yang, X. Zhang, H. Liu, and B. Tao, "Multi-Feature Input Deep Forest for EEG-Based Emotion Recognition," *Front Neurorobot*, vol. 14, Jan. 2021, https://doi.org//10.3389/fnbot.2020.617531

[16] C. Filippini et al., "Automated Affective Computing Based on Bio-Signals Analysis and Deep Learning Approach," *Sensors*, vol. 22, no. 5, Mar. 2022, https://doi.org//10.3390/s22051789

[17] Kusumaningrum, T. D., Akhmad Faqih, and Benyamin Kusumoputro. "Emotion recognition based on DEAP database using EEG time-frequency features and machine learning methods." In *Journal of Physics: Conference Series*, vol. 1501, no. 1, p. 012020. IOP Publishing, 2020.https://doi.org//10.1088/1742-6596/1501/1/012020

[18] Jolly, Mili, Pranjal Gupta, Sanskriti Bansal, Garima Gupta, and Ruchi Goel. "Machine learning based speech emotion recognition in hindi audio." In *2023 7th International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 1382-1388. IEEE, 2023. https://doi.org//10.1109/ICOEI56765.2023.10125902

[19] Cheng, Wen Xin, Ruobin Gao, P. N. Suganthan, and Kum Fai Yuen. "EEG-based emotion recognition using random Convolutional Neural Networks." *Engineering applications of artificial intelligence* 116 (2022): 105349. https://doi.org//10.1016/j.engappai.2022.105349

[20] IEEE Computational Intelligence Society, International Neural Network Society, Institute of Electrical and Electronics Engineers, and *IEEE World Congress on Computational Intelligence (2020 : Online)*, *2020 International Joint Conference on Neural Networks (IJCNN) : 2020 conference proceedings*.

[21] Sharma, Tanya, Manoj Diwakar, Prabhishek Singh, Chandrakala Arya, Sumita Lamba, and Pramod Kumar. "A review on EEG based Emotion Analysis using Machine Learning approaches." In *2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, pp. 1-6. IEEE, 2021. https://doi.org//10.1109/UPCON52273.2021.9667588

[22] F. Al Machot, A. Elmachot, M. Ali, E. Al Machot, and K. Kyamakya, "A deep-learning model for subject-independent human emotion recognition using electrodermal activity sensors," *Sensors (Switzerland),* vol. 19, no. 7, Apr. 2019, https://doi.org//10.3390/s19071659

[23] Nakisa, Bahareh, Mohammad Naim Rastgoo, Andry Rakotonirainy, Frederic Maire, and Vinod Chandran. "Automatic emotion recognition using temporal multimodal deep learning." *IEEE Access* 8 (2020): 225463-225474. https://doi.org//10.1109/ACCESS.2020.3027026

[24] S. Rafique, N. Kanwal, M. S. Ansari, M. Asghar, and Z. Akhtar, "Deep Learning based Emotion Classification with Temporal Pupillometry Sequences," *in International Conference on Electrical, Computer, and Energy Technologies, ICECET 2021*, Institute of Electrical and Electronics Engineers Inc., 2021. https://doi.org//10.1109/ICECET52533.2021.9698663

[25] L. Yang and J. Liu, "EEG-Based Emotion Recognition Using Temporal Convolutional Network.", https://doi.org//10.1109/DDCLS.2019.8908839

[26] Singh, Gurwinder, Kanupriya Verma, Neha Sharma, Ashok Kumar, and Archana Mantri. "Emotion recognition using deep convolutional neural network on temporal representations of physiological signals." In *2020 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT)*, pp. 1-6. IEEE, 2020. https://doi.org//10.1109/ICMLANT50963.2020.9355990

[27] Suzuki, Kei, Tipporn Laohakangvalvit, Ryota Matsubara, and Midori Sugaya. "Constructing an emotion estimation model based on EEG/HRV indexes using feature extraction and feature selection algorithms." *Sensors* 21, no. 9 (2021): 2910. https://doi.org//10.3390/s21092910

[28] Y. Lim, K.-W. Ng, P. Naveen, and S.-C. Haw, "Emotion Recognition by Facial Expression and Voice: Review and Analysis," *Journal of Informatics and Web Engineering*, vol. 1, no. 2, pp. 45–54, Sep. 2022, https://doi.org//10.33093/jiwe.2022.1.2.4

[29] S. Oh, J. Y. Lee, and D. K. Kim, "The design of CNN architectures for optimal six basic emotion classification using multiple physiological signals," *Sensors (Switzerland)*, vol. 20, no. 3, Feb. 2020, https://doi.org//10.3390/s20030866

[30] Cimtay, Yucel, and Erhan Ekmekcioglu. "Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset EEG emotion recognition." *Sensors* 20, no. 7 (2020): 2034. https://doi.org//10.3390/s20072034

[31] Dzieżyc, Maciej, Martin Gjoreski, Przemysław Kazienko, Stanisław Saganowski, and Matjaž Gams. "Can we ditch feature engineering? end-to-end deep learning for affect recognition from physiological sensor data." *Sensors* 20, no. 22 (2020): 6535.https://doi.org//10.3390/s20226535

[32] Algarni, Mona, Faisal Saeed, Tawfik Al-Hadhrami, Fahad Ghabban, and Mohammed Al-Sarem. "Deep learning-based approach for emotion recognition using electroencephalography (EEG) signals using bi-directional long short-term memory (Bi-LSTM)." *Sensors* 22, no. 8 (2022): 2976. https://doi.org//10.3390/s22082976

[33] S. Praveenkumar and T. Karthick, "Human Stress Recognition by Correlating Vision and EEG Data," Computer Systems Science and Engineering, vol. 45, no. 3, pp. 2417–2433, 2023, https://doi.org//10.32604/csse.2023.032480

[34] Mahmud, Wan Mahani Hafizah Wan, et al. "Optical Coherence Tomography Image Analysis for Detection of Alzheimer's Disease: A Comprehensive Structured Review." *Journal of Advanced Research in Applied Sciences and Engineering Technology (2024):* 41-57. https://doi.org/10.37934/araset.61.2.4157

[35] Subramanian, Ramanathan, Julia Wache, Mojtaba Khomami Abadi, Radu L. Vieriu, Stefan Winkler, and Nicu Sebe. "ASCERTAIN: Emotion and personality recognition using commercial sensors." *IEEE Transactions on Affective Computing* 9, no. 2 (2016): 147-160.https://doi.org//10.1109/TAFFC.2016.2625250