E-ISSN: 2583-9667
Indexed Journal
Peer Reviewed Journal
https://multiresearchjournal.theviews.in

Received: 07-05-2025
Accepted: 16-06-2025
Published: 08-07-2025

INTERNATIONAL JOURNAL OF ADVANCE RESEARCH IN MULTIDISCIPLINARY

Volume 3; Issue 3; 2025; Page No. 278-285

Algorithm for multimodal emotion recognition using deep learning networks

[1]Amit Ghosh and [2]Dr. Santanu Sikdar

[1]Research Scholar, P.K University, Shivpuri, Madhya Pradesh, India
[2]Professor, P.K University, Shivpuri, Madhya Pradesh, India

DOI: https://doi.org/10.5281/zenodo.17295787

Corresponding Author: Amit Ghosh

Abstract

Understanding and detecting human emotions by combining data from several sources, such as speech, text, facial expressions, and physiological signs, is known as multimodal emotion identification. Ubiquitous Computing describes how computers and their apps have become an integral part of modern living. Users' interactions with computers are increasingly resembling those between humans. This study presents a novel deep-learning method based on the BROA to facilitate face emotion identification using several data modalities. The BROA optimization method integrates the Bat method (BA) with the Rider Optimization Algorithm (ROA). This integration seeks to improve the precision of emotion identification by adeptly using facial pictures, EEG data, and physiological signals as components of the multimodal input.

Keywords: BROA, Physiological Signs, ROA, Emotional intelligence

Introduction

Emotions are multi-faceted mental and physical states that may be set off by several internal and external elements. They consist of a broad range of subjective experiences that might impact one's ideas, actions, and physiological reactions, including emotions, moods, and affective states. As a result, the scientific community has put out an increasing number of definitions for the concept of emotion, despite the fact that it is a widely used and seemingly simple phrase.

Ideas, the intensity of positive or negative experiences, and behavioral responses all impact an individual's emotional state. Human emotion does not have a single, widely recognized definition at this time. Over the last 20 years, researchers in domains as diverse as computer science, neuroscience, and psychology have made great strides in understanding human emotion. The literature is subjective when it comes to defining human feeling. Emotions may be better understood by using theories from both cognitive neuroscience and physiological research. According to the physiological view, feelings originate in the human body. It is possible that the alteration in physiological reaction is due to the outside incident. A common example is the widespread fear of snakes. According to cognitive neuro-science theory, it is the brain's own processes that are responsible for producing an emotional response. The prefrontal cortex is responsible for controlling emotions and making decisions, as well as storing information about oneself. Thoughts and other mental pursuits are linked to emotions. As an example, seeing the snake makes individuals feel threatened and causes them to become scared. Emotions are defined as changes in physiological and cognitive reactions to stimuli, such as pleasure or pain.

A new topic of emotional computing, multimodal emotion detection is only starting to get off the ground. Combinations of auditory, visual, and physiological inputs are known as multimodal signals. The video and audio signals show how people react when they're feeling a certain mood via their facial expressions and words. When a person experiences an emotion, their physiological signals reveal it via changes in their temperature, breathing rate, electrical activity in the brain, and muscular current. In order to differentiate between different human emotions, researchers are working on a multimodal emotion detection system, which is a significant pattern recognition challenge. To improve emotion identification, it incorporates data from

people's face expressions, voice, and bodily symptoms. Human facial expressions and audio data are taken into account for multimodal emotion identification in this thesis. Consideration of speech and facial expressions for emotion identification is warranted for a number of reasons.

The impetus for using multimodal emotion detection via deep learning networks and fusion approaches is to develop systems that are more precise, resilient, and contextually cognizant in interpreting and reacting to human emotions. Human emotions are expressed via facial expressions, voice tones, gestures, body movements, and posture. An automated system proficient in recognizing these emotions might transform several fields, such as video games, human-computer interaction, robotics, instructional software, animations, car safety, and affective computing. Consequently, creating a dependable, real-time emotion identification system is an essential objective, with its many applications necessitating comprehensive investigation. A hopeful consequence is the development of increasingly advanced robots capable of comprehending human emotions, resulting in substantial progress in this domain

Literature Review

H. Wang *et al*. (2020) [1] an emotion database that was created using electroencephalogram data. In accordance with 10-20 worldwide electrode placement methods, A 128-channel Geodesic Sensor Net system running at 250 Hz was used to capture the EEG data. 16 participants with an average age of 22.27 years and a standard deviation of 1.37 were involved in the study; half of them were men. Every participant saw 12 appropriate video segments, each lasting about 4 minutes, designed to evoke a range of emotions, from happiness to sadness to neutrality. Twelve trials were used to capture good, negative, and neutral feelings throughout three sessions and four trials per session. In order to determine how each participant really filled out the EEG recording form, a self-assessment feedback form was filled out after every session. Appropriate feelings were induced by allowing enough time to relax after each session.

T. B. Alakus *et al*. (2020) [2] 'GAMEEMO', an emotion recognition system's EEG-based database, was launched. Participants had their electroencephalograms (EEGs) monitored as they played video games. An EEG data from 28 different people was recorded using a 14-channel EMOTIV EPOC+ EEG equipment. 2048 Hz was the first sampling rate, and it was subsequently down sampled to 128 Hz. During the EEG recording, each patient played one of four 5-minute computer games: bored, calm, terror, and fun. Using a 5th order sinc filter, we were able to eliminate artifacts caused by the user's head, arm, and hand movements.

T. Song *et al*. (2019) [3] the "MPED: A Multi-Modal Physiological Emotion Database for Discrete Emotion Recognition" framework. The 62-channel ESI NeuroScan EEG equipment was used for EEG recording at a 1000 Hz sampling rate, and MPED adhered to the standards of the worldwide 10–20 electrode arrangement. The BIOPAC gadget captured RSP, ECG, and GSR at a sampling rate of 250 Hz. Twenty-three healthy people without a history of neurological or mental disorder were enrolled in the study. Their mean age was 20, and they ranged in age from 18 to 25. The experiment was split into two portions with a minimum 24-hour delay between each to provide participants ample time to relax, since prolonged studies may lead to participant tiredness. A 120-second resting state, fourteen trials, and a self-assessment evaluation made up each independent experiment. While Participants were told to close their eyes and relax during the 120-second resting condition; during this time, psychological signals were also recorded. Each trial began with a 10-second countdown, followed by a 30-second rest then the 2.5–5-minute Chinese video clip. The clips from the video were shown in a random sequence and had all the essential elements to evoke the desired feeling.

A. Baghdadi *et al*. (2019) [4] presented "DASPS: A Database for Anxious States based on a Psychological Stimulation." Fifteen male and thirteen female healthy volunteers had their electroencephalograms (EEGs) recorded. Thirteen women and ten men participated, having an average age of almost thirty years. The participants were made to feel anxious by the challenging math's issue. The Emotive EPOC 14 channel was used to capture the EEG data at a sampling rate of 128 Hz in compliance with the global 10-20 electrode placement scheme. Every one trial's recording was split into individual 1-minute segments from each 6-minute recording.

M. Bachmann *et al*. (2017) [5] compiled a list of depressed individuals. The trials included 34 females (mean age-39 years), 17 of whom were depressed patients and 17 of whom served as controls. The participants were chosen from an inpatient ward of a hospital and all had major depressive illness. The criteria established by ICD-10 were used to identify subjects who did not have psychotic depressive disorder. EEG data was collected from 9 a.m. till noon at 30-minute intervals. The electroencephalogram (EEG) was captured using a Cadwell Easy II 400 Hz sample rate, in accordance with the global 10-20 electrode placement scheme. Eye movements and muscular activity were identified and removed as artifacts by visual inspection. Cutoff frequencies for the digital filtering of the EEG signals ranged from half a hertz to forty hertz.

Materials and Methods

This research employs multimodal signals, which include physiological data, face photos, and EEG. First, to eliminate any undesired artifacts, the facial image is pre-processed. After that, features are extracted from the processed picture using LDTOOP, a combination of LOOP and LDTP. Similarly, features like wavelet coefficients and spectrum parameters including PSD, spectral flux, tonal power ratio, spectral skewness, and spectral centroids are used to get the EEG and physiological signals ready for analysis. extracted. Afterwards, a DBN classifier processes the characteristics obtained from the face picture.

Gathering of input video: An important aspect of emotional computing systems is emotion recognition. A person's emotional state is multi-faceted, including not just their thoughts and actions but also their physiological responses to external stimuli. Physiological signals, video, and electroencephalogram (EEG) data are common components of facial emotion detection systems. When training and testing face emotion detection algorithms, one popular dataset is the DEAP dataset.

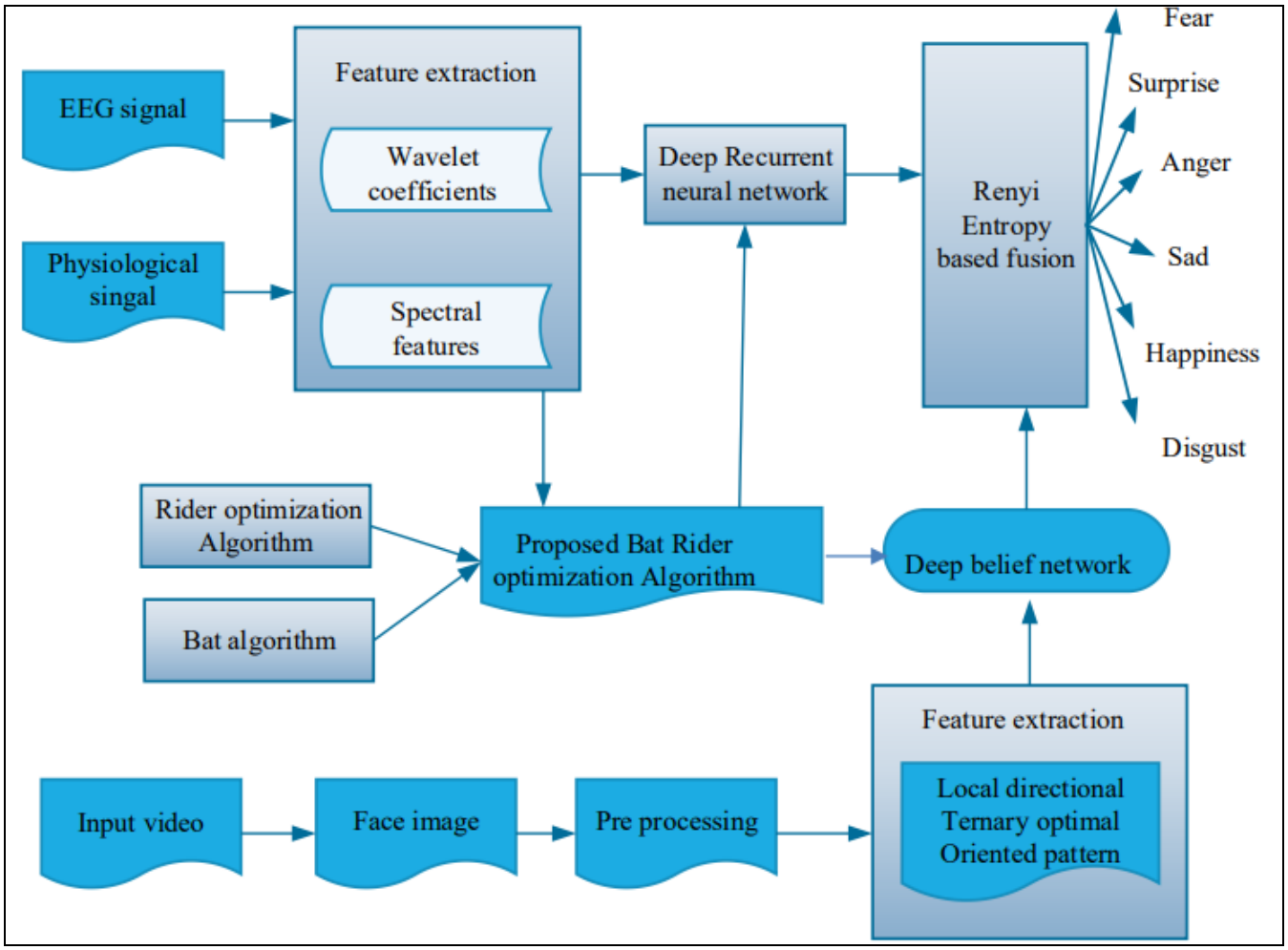

Fig 1: Proposed BROA-based Deep NN schematic view

Image extraction for the face: Here we'll assume a dataset A and use the variable to indicate the number of movies included inside it.

$$H = \{A_1, A_2 \ldots A_i \ldots . A_n\} \qquad (1)$$

Here, H stands for the dataset and n for the number of movies. Face shots are included in each film from different angles. In dataset H, the ith movie contains facial pictures, which may be represented as follows:

$$A_i = \{X_1, X_2 \ldots X_i \ldots . X_n\} \qquad (2)$$

The ith facial picture from the ith video is represented by A in this context. There is a unique location for each individual's face picture, and each face image belongs to a certain person. The pre-processed result is the result of taking the face picture from the dataset and running it through a pre-processing step.

Image pre-processing for faces: To ensure that the input face picture is of high enough quality and suitable for further analysis, the pre-processing module is an essential part of face image analysis. The main goal of the several processes that make up this module is to remove noise and external artifacts. After importing the face picture into the pre-processing module, the image undergoes noise reduction.

DBN-based facial emotion recognition

In order to categorize emotions from face images, the DBN classifier is used. The DBN classifier uses these characteristics as inputs to recognize face emotions. With connections created between visible and buried neurons, the DBN architecture consists of two RBM layers followed by one MLP layer. Importantly, the next layer takes as input the output of the previous layer.

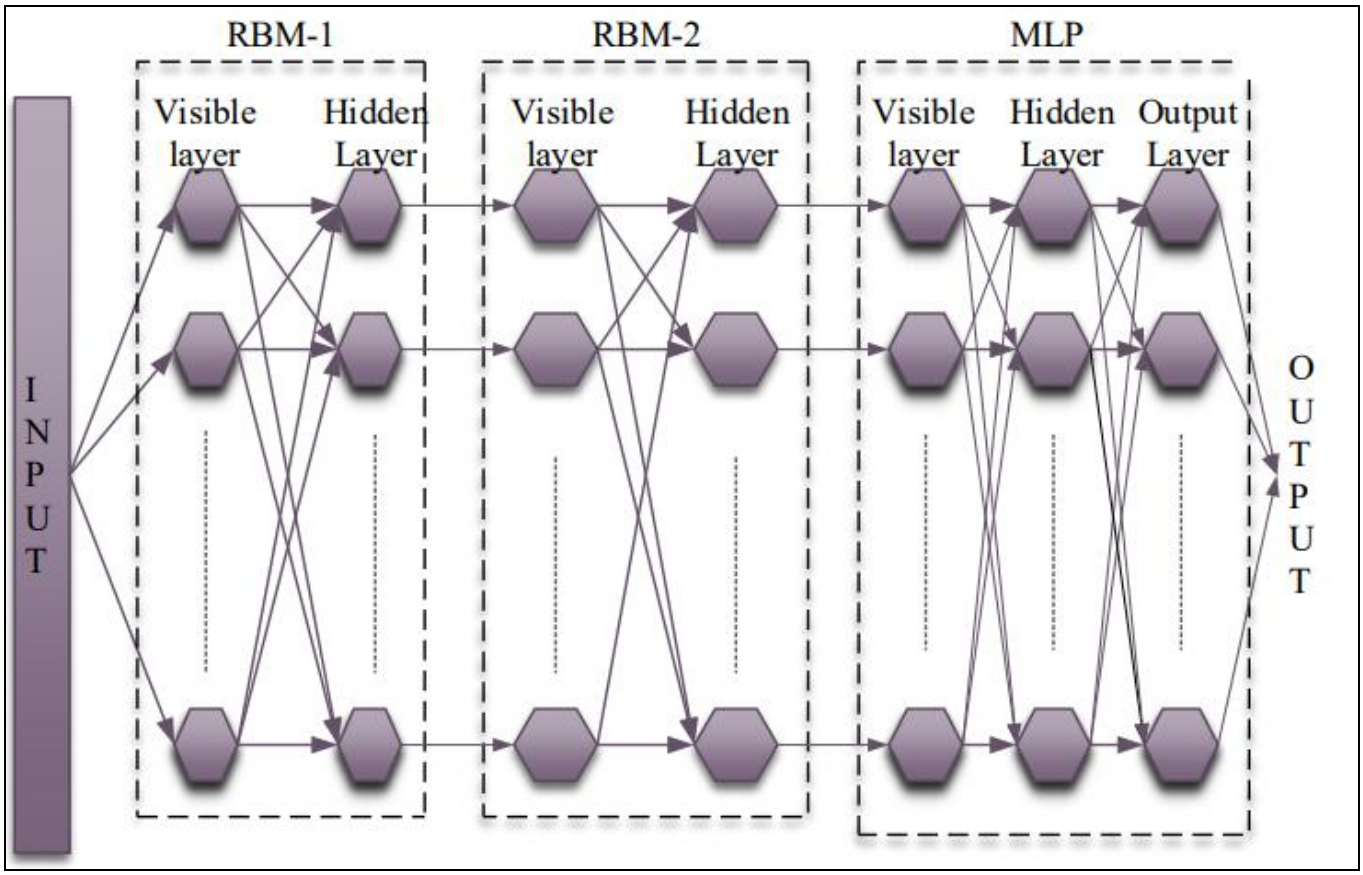

Fig 2: Recognizing emotions with DBN

The BROA-DNN was adopted as the basic deep learning framework in this study. Practically, the architecture was incorporated into the experimental pipeline by initially implementing the preprocessing steps that have been discussed above (resizing inputs, normalization, and augmentation). The data underwent processing and then it was ingested into the BROA-DNN where the layered nature of the model model made features learn in hierarchical order- starting with low-level edge and texture representation, as well as, high-level semantic representations of expressions.

Results

Analysis with Different Hidden Neurons

Accuracy for Hidden Neuron: Depending on the quantity of hidden neurons and training data, the suggested BROA-based deep learning algorithm showed varied degrees of accuracy. With 60% of the data used for training, the accuracy improved with more hidden neurons, ranging from 0.8467 to 0.8978. Accuracy, which ranged from 0.8654 to 0.9165, remained good at 70% of the training data. Accuracy peaked at 0.9356 with 100 hidden neurons, and it improved at 90% of training data. Thanks to improvements in accuracy brought about by the selection of hidden neurons and an increase in training data, these results demonstrate how successful the approach is in emotion recognition.

Table 1: Hidden Neuron Accuracy

Training %	Hidden Neurons	Accuracy
60%	25	0.8467
60%	50	0.8654
60%	75	0.8865
60%	100	0.8978
70%	25	0.8654
70%	50	0.8865
70%	75	0.8967
70%	100	0.9165
80%	25	0.8867
80%	50	0.8967
80%	75	0.9087
80%	100	0.9256

With a variety of training data percentages and hidden neuron counts shown in this table, the accuracy of the suggested BROA-based deep learning approach may be easily grasped. As a result, you can see how the accuracy varies in various settings.

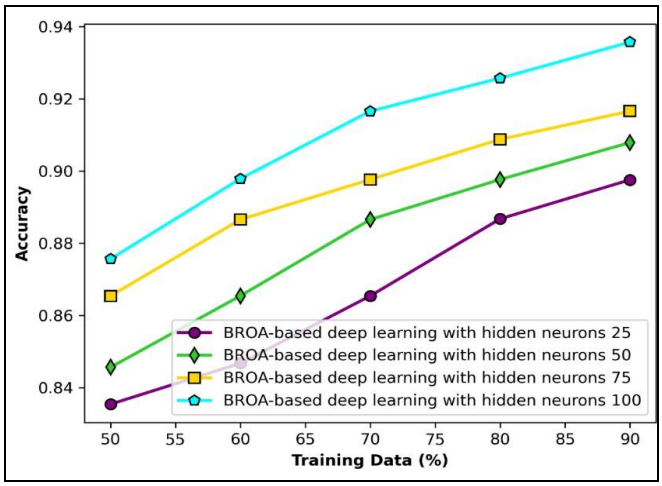

Fig 3: Accuracy for Hidden Neuron with different training values

False Acceptance Rate (FAR) for Hidden Neuron: Using FAR, we were able to identify some interesting patterns in our proposed BROA-based deep learning approach. The FAR drops with increasing numbers of hidden neurons with 80% training data, reaching a minimum of 0.165 with 100 hidden neurons. With an astounding FAR of 0.146 and 100 hidden neurons, the model's capacity to decrease erroneous acceptance is further shown by further gains at 90% of training data. This shows that the model is good at reducing false positives considering the quantity of buried neurons and the dimensions of the training dataset grows.

Table 2: Hidden Neuron Progression

Training %	Hidden Neurons	FAR
70%	25	0.257
70%	50	0.226
70%	75	0.215
70%	100	0.175
80%	25	0.235
80%	50	0.217
80%	75	0.208
80%	100	0.165
90%	25	0.222
90%	50	0.208
90%	75	0.187
90%	100	0.146

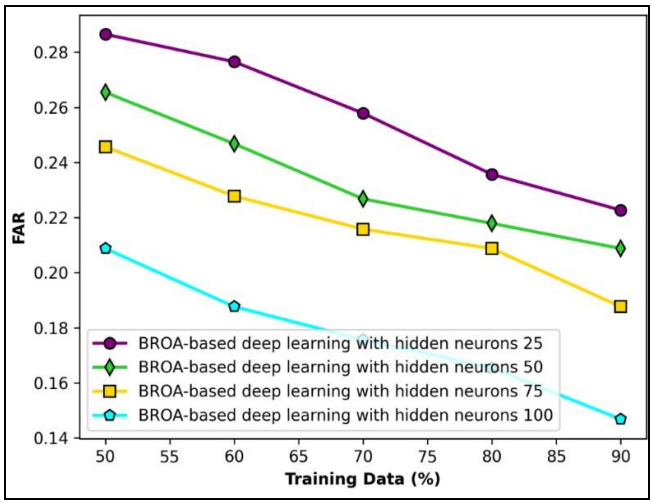

Fig 4: FAR for Hidden Neuron with different training values

At various training data percentages and with varied numbers of hidden neurons, the proposed BROA-based deep learning method's False Acceptance Rate (FAR) is clearly shown in this table. By doing so, the method's performance under various scenarios may be evaluated.

False Rejection Rate (FRR) for Hidden Neuron

With varied numbers of hidden neurons and training data percentages, FRR is seen in the suggested deep learning method based on BROA. The FRR values declined as the number of hidden neurons increased at 70% training data; the lowest FRR recorded at 0.193 was reached with 100 hidden neurons. At 80% of the training data, the same patterns were visible; adding more hidden neurons to the model always resulted in an improved FRR. At last, the FRR dropped, hitting a low of 0.176 with 100 hidden neurons at 90% of the training data. This shows that the model can decrease the number of erroneous rejections with more hidden neurons and more training datasets.

Table 3: Hidden Neuron FRR

Training %	Hidden Neurons	FRR
70%	25	0.240
70%	50	0.238
70%	75	0.228
70%	100	0.193
80%	25	0.238
80%	50	0.234
80%	75	0.224
80%	100	0.183
90%	25	0.234
90%	50	0.225
90%	75	0.215
90%	100	0.176

At various training data percentages and with varied numbers of hidden neurons, the proposed BROA-based deep learning method's False Rejection Rate (FRR) is clearly seen in this table. This is useful for evaluating the method's robustness under various scenarios, especially with regard to its capacity to reduce the occurrence of erroneous rejections.

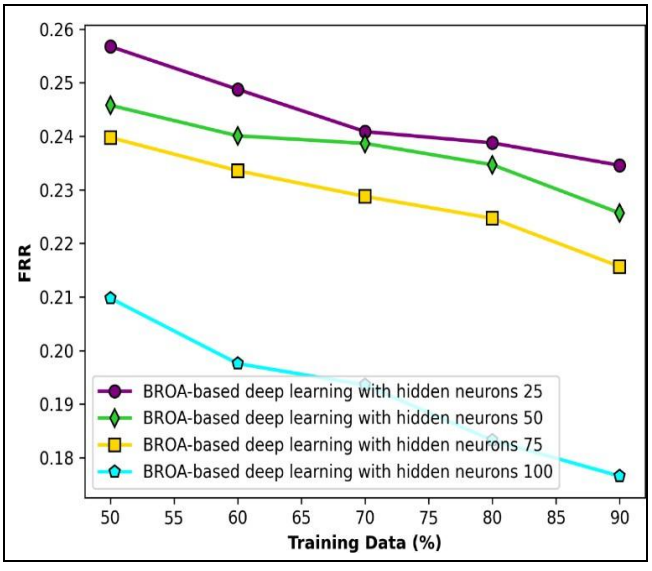

Fig 5: FRR for Hidden Neuron with different training value

Analysis with Different Rates of Learning

An examination of the rate of learning looks at the speed with which a machine learning model reaches a solution during training. The model's performance is usually evaluated across several learning rates, which govern the step size in gradient-based optimization techniques. When the learning rate is too high, overshooting is possible; when it's too low, convergence is sluggish or the algorithm becomes trapped in local minima.

Accuracy for Rate of Learning

A BROA-based deep learning model was capable to attain a range of accuracy values at a 70% training data level, depending on the learning rate: 0.05 with an accuracy of 0.8367, 0.1 with an accuracy of 0.8465, 0.15 with an accuracy of 0.8577, and 0.2 with an accuracy of 0.8976. The model's accuracy fluctuated with the same set of learning rates at 80% training data: 0.05 with 0.8433 accuracies, 0.1 with 0.8545 accuracies, 0.15 with 0.8755 accuracy, and 0.2 with 0.9076 accuracy. Finding the sweet spot for learning rates in emotion detection is crucial, as this data shows.

Table 4: Learning Rate Accuracy

Training %	Rate of Learning	Accuracy
70%	0.05	0.8367
70%	0.1	0.8465
70%	0.15	0.8577
70%	0.2	0.8976
80%	0.05	0.8433
80%	0.1	0.8545
80%	0.15	0.8755
80%	0.2	0.9076

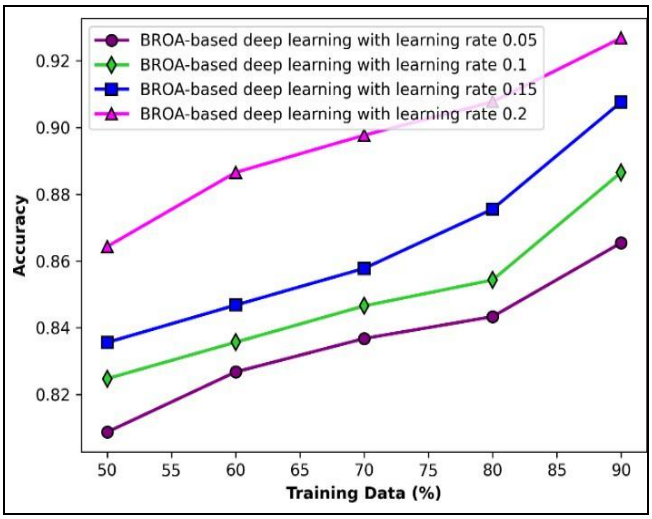

Fig 6: Accuracy for Rate of Learning with different training values

SAR for Learning Rate

The proposed BROA-based deep learning approach showed various FAR at different learning rates at an 80% training data level: 0.05 (0.0581), 0.1 (0.0644), 0.15 (0.0772), and 0.2 (0.1122). The models' FARs were also quite constant at 90% of training data for the same learning rates: 0.05 (0.0557), 0.1 (0.0604), 0.15 (0.0604), and 0.2 (0.0609). This study shows that choosing the right learning rate is crucial for improving the model's accuracy on emotion identification tasks, as various rates provide diverse results.

Table 5: Results for Rate of Learning

Training %	Rate of Learning	FAR
80%	0.05	0.0581
80%	0.1	0.0644
80%	0.15	0.0772
80%	0.2	0.1122
90%	0.05	0.0557
90%	0.1	0.0604
90%	0.15	0.0604
90%	0.2	0.0609

Various learning rates and training data percentages are shown in this table, along with the suggested BROA-based deep learning's accuracy algorithm. Taking into account the learning rate and the training data used, it aids in evaluating the model's performance under different scenarios.

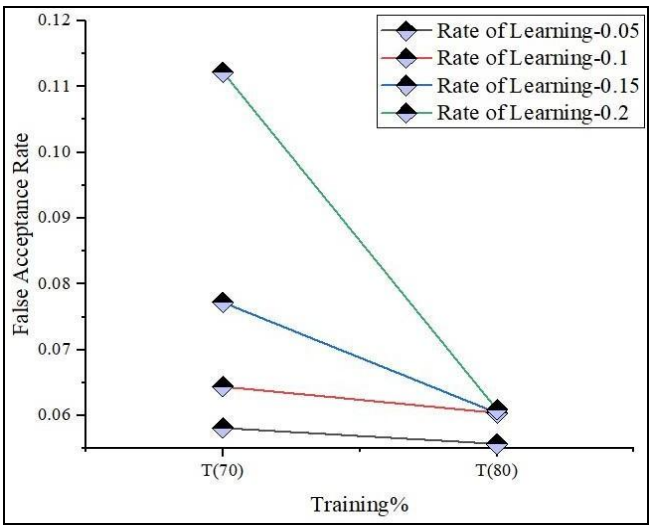

Fig 7: FAR for Rate of Learning with different training values

FRR for Rate of Learning

With 70% training data, the suggested BROA-based deep learning method's FRR changed from 0.05 (0.1987) to 0.2 (0.1466) as a function of learning rate. At an 80% level of training data, the model's FRR varied at the same learning rates: 0.05 (0.1876), 0.1 (0.1756), 0.15 (0.1578), and 0.2 (0.1356). This demonstrates how choosing a learning rate affects the model's performance in emotion identification, as different rates provide different levels of accuracy.

Table 6: Rate of Learning FRR

Training %	Rate of Learning	FRR
70%	0.05	0.1987
70%	0.1	0.186
70%	0.15	0.1765
70%	0.2	0.1466
80%	0.05	0.1876
80%	0.1	0.1756
80%	0.15	0.1578
80%	0.2	0.1356

Various learning rates and training data percentages are shown in this table, along with the suggested BROA-based deep learning algorithm's accuracy. By taking the learning rate and training data into account, it aids in evaluating the model's performance under different scenarios.

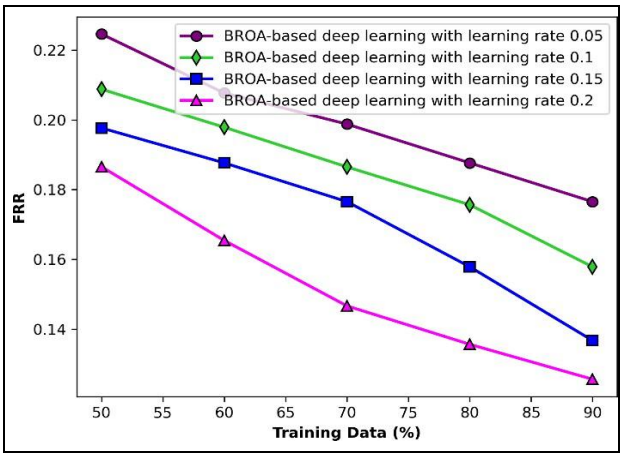

Fig 8: FRR for Rate of Learning with different training values

Comparison With Existing Methods

We compare the performance of the suggested technique in comparison to a number of existing methods, such as Deep Multimodal Attentive Fusion (MAF), LSTM with RNN, NN, and Attention-based Bidirectional LSTM with RNN. A thorough comparison is conducted in this study to measure the performance of the suggested strategy compared to these recognized approaches.

Accuracy of the Suggested BROA-Deep NN Compared to Current Techniques

Table 7: Comparison of Accuracy Values

Methods	Accuracy
BLSTM-RNN	0.8375
DMAF	0.8435
LSTM-RNN	0.8765
ANN	0.9045
Proposed BROA-Deep NN	0.9356

Various approaches to emotion identification using multimodal signals have shown the following accuracy results: The following results were obtained: 0.8375 for BLSTM-RNN, 0.8435 for DMAF, 0.8765 for LSTM-RNN, and 0.9045 for ANN. With an accuracy of 0.9356, the suggested BROA-Deep NN demonstrated its ability in reliably distinguishing emotions from multiple data sources, surpassing all existing approaches. The findings show that the BROA-Deep NN approach is the best one.

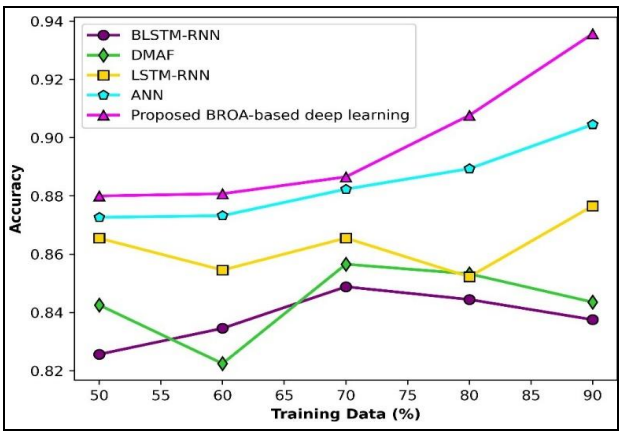

Fig 9: Comparative Analysis for Accuracy

FAR of Proposed BROA-Deep NN with the Existing Methods: Various approaches' False Acceptance Rates (FARs) are evaluated in the performance evaluation.

Table 8: Values for FAR compared to one another

Methods FAR	
LSTM-RNN	0.2376
DMAF	0.2154
BLSTM-RNN	0.2175
ANN	0.2046
Proposed BROA-deepNN	0.1756

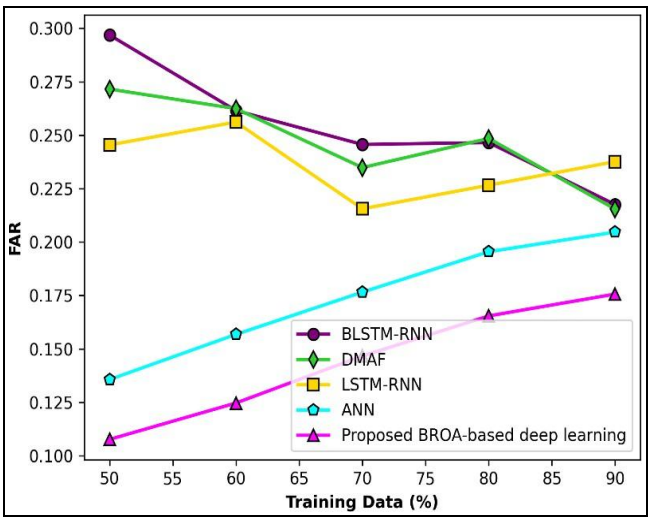

Fig 10: Comparative Analysis for FAR

It is evident from the comparison that LSTM-RNN had a lower FAR of 0.2376. The FAR that ANN reported was a pitiful 0.2046. A modest FAR of 0.2175 was attained using BLSTM-RNN. DMAF showed a somewhat better FAR of 0.2154. With a lowest FAR of 0.1756, the suggested BROA-Deep NN technique significantly outperformed the others. The accuracy of each method in reducing false alarms in emotion identification tasks is shown by these FAR values. Reducing false detections and improving overall accuracy of emotion identification from varied multimodal data sources is highlighted by the outstanding performance of the proposed BROA-Deep NN, which is characterized by the lowest FAR.

FRR of the Proposed BROA-Deep NN with the Existing Methods: With a FRR of 0.1876, the BLSTM-RNN approach proved to be rather successful for this specific assignment. DMAF showed a competitive approach with a slightly lower FRR of 0.1654. Among the studied approaches, LSTM-RNN stood out with a FRR of 0.1577, indicating greater performance. In contrast, the ANN method's rather substantial FRR of 0.1367 suggests it may not be the best fit for the job compared to the other approaches. With the lowest FRR of 0.1267, the Proposed BROA-deep NN strategy surpassed all other approaches, suggesting it might be an efficient solution for the issue at hand and deserving of more investigation in future research efforts. Using these FRR values as quantitative measurements, we can compare the performance of different methods; lower values indicate that the methods are more suited to the job at hand.

Table 9: Values for FRR compared

Methods FRR	
BLSTM-RNN	0.1876
DMAF	0.1654
LSTM-RNN	0.1577
ANN	0.1367
Proposed BROA-Deep NN	0.1267

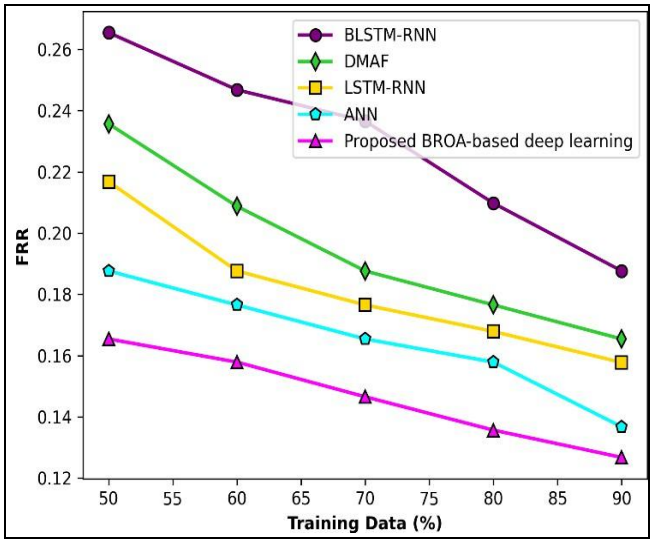

Fig 11: Comparative Analysis for FRR

Conclusion

To make the most of the characteristics that were collected and to enable accurate categorization, a deep learning classifier is used. Facial expression identification is one area where deep learning models shine because they can automatically extract hierarchical representations from complex data. Here, we use the BROA-based deep learning classifier, which optimizes the classifier's parameters by combining the advantages of BA and ROA. The study findings show that the deep learning strategy that was suggested using BROA is successful. With an astounding maximum accuracy of 0.9356, it accomplishes an extraordinary degree of performance. In addition, it reduces the FRR to an even lower minimum of 0.1267 and the FAR to a remarkable minimum of 0.1756. The method's ability to accurately recognize emotions has great practical implications in domains such as psychology, human-computer interaction, and emotional computing. In the end, the BROA-based deep learning methodology exhibits significant potential for advancements in the fields of theory and practice multimodal emotion identification.

References

1. Wang H, Wu X, Yao L. Identifying cortical brain directed connectivity networks from high-density EEG for emotion recognition. IEEE Transactions on Affective Computing. 2020.
2. Alakus TB, Gonen M, Turkoglu I. Database for an emotion recognition system based on EEG signals and various computer games – GAMEEMO. Biomedical Signal Processing and Control. 2020;60:101951.
3. Song T, Zheng W, Lu C, Zong Y, Zhang X, Cui Z. MPED: A multi-modal physiological emotion database for discrete emotion recognition. IEEE Access. 2019;7:12177–12191.
4. Baghdadi Y, Aribi R, Fourati R, Halouani N, Siarry P, Alimi AM. DASPS: A database for anxious states based on a psychological stimulation. arXiv preprint arXiv:1901.02942. 2019.
5. Bachmann M, Lass J, Hinrikus H. Single channel EEG analysis for detection of depression. Biomedical Signal Processing and Control. 2017;31:391–397.
6. Katsigiannis S, Ramzan N. DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. IEEE Journal of Biomedical and Health Informatics. 2017;22(1):98–107.
7. Zheng W-L, Zhu J-Y, Lu B-L. Identifying stable patterns over time for emotion recognition from EEG. IEEE Transactions on Affective Computing. 2017;10(3):417–429.
8. Wei L-L, Chen Z-Z, Song Z-Z, Lou X-G, Li D-D. EEG-based emotion recognition using simple recurrent units network and ensemble learning. Biomedical Signal Processing and Control. 2020;58:101756.
9. Sharma R, Pachori RB, Sircar P. Automated emotion recognition based on higher order statistics and deep learning algorithm. Biomedical Signal Processing and Control. 2020;58:101867.
10. Chakladar D, Dey S, Roy PP, Dogra DP. EEG-based mental workload estimation using deep BLSTM-LSTM network and evolutionary algorithm. Biomedical Signal Processing and Control. 2020;60:101989.
11. Cimtay Y, Ekmekcioglu E. Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset EEG emotion recognition. Sensors. 2020;20(7):2034.
12. Pandey P, Seeja K. Subject independent emotion recognition from EEG using VMD and deep learning. Journal of King Saud University – Computer and Information Sciences. 2019.
13. Ullah H, Uzair M, Mahmood A, Ullah M, Khan SD, Cheikh FA. Internal emotion classification using EEG signal with sparse discriminative ensemble. IEEE Access. 2019;7:40144–40153.
14. Wang Z-M, Hu S-Y, Song H. Channel selection method for EEG emotion recognition using normalized mutual information. IEEE Access. 2019;7:143303–143311.
15. Li Y, Zheng W, Wang L, Zong Y, Cui Z. From regional to global brain: a novel hierarchical spatial-temporal neural network model for EEG emotion recognition. IEEE Transactions on Affective Computing. 2019.