Emotion classification on EEG brainwave dataset

Ashish Sah[1]

IIT Gandhinagar
sahashish@iitgn.ac.in

Abstract. Emotion plays a crucial role in human behavior and decision-making, and emotion classification finds its application in various fields such as criminal detection, security services and many more. However, there is no one single method/algorithm which can always be best suited for different varieties of data. This project explores the performance of various machine learning and deep learning methods to classify emotional experiences based on EEG brainwave data. We explored some algorithms such as Random Forest, Logistic Regression, Artificial neural network, Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU), for the efficient classification of emotional state. The dataset used for this experiment consists of EEG signals recorded from individuals while experiencing different emotional states, which were then labeled accordingly. The obtained result shows that most of the deep learning models performed very well where the LSTM model was reported with an accuracy of 98.2%. Similarly GRU's accuracy was reported to be 96.02%. In contrast, the Random Forest and logistic regression model resulted into score of 92% and 77.02% respectively.

Keywords: EEG · Emotion · Deep learning · LSTM

Overleaf: https://www.overleaf.com/read/fxsdrdptymgm

1 Introduction

Emotions are an essential aspect of human behavior and play a crucial role in our everyday lives. They influence our decision-making, perception, and cognition, and are closely linked to our physical and mental well-being. The ability to accurately detect and classify emotions can provide valuable insights into human behavior and has numerous application in various artificial intelligence based tools such as patient monitoring, criminal detection, disabled assistance, security services, robotics, communication and many other.
Electroencephalography (EEG) is a non-invasive method for measuring the electrical activity of the brain and has been widely used in research to investigate neural correlates of emotion. EEG data provides valuable information about the temporal dynamics of brain activity associated with emotional processing, making it a promising approach for emotion detection and classification.
In recent years, machine learning techniques have shown promising results in

the analysis of EEG data for emotion classification. Various algorithms have been explored, including support vector machines, random forests, and neural networks. However, there is still a need for further investigation and comparison of other algorithms to determine the most effective approach for emotion classification using EEG data. This project aims to fill this gap by exploring the performance of multiple machine learning algorithms such as LSTM, GRU, Random Forest, SVM[6] along with various feature extraction algorithms such as PCA, on a dataset[2] of EEG data to classify emotions. The project evaluates the accuracy, precision, and recall of these algorithms and compare their performance to determine the most effective approach for emotion detection and classification using EEG data.

2 Related Work

Mental emotional sentiment classification is an emerging field that aims to classify mental and emotional states using EEG signals. The use of EEG-based Brain Machine Interfaces (BMI) has shown promising results in this area, enabling the development of intelligent systems that can interpret and respond to human mental states in real-time.
Jordan J. Bird, in their paper "Mental emotional sentiment classification with an EEG-based brain-machine interface,"[2] proposed a machine learning-based approach for mental and emotional sentiment classification using EEG signals. The paper outlines a method that involves extracting features from EEG signals using time-frequency analysis and training a Support Vector Machine (SVM) classifier on these features. The authors report promising results in the classification of mental and emotional states, highlighting the potential of EEG-based BMIs for the development of intelligent systems that can interpret and respond to human mental states.

In a similar vein, the paper "A deep evolutionary approach to bioinspired classifier optimization for brain-machine interaction"[1] by Jordan J. Bird proposes a deep evolutionary approach for bioinspired classifier optimization in BMIs. The authors used a deep learning-based approach that involves evolving a neural network for EEG signal classification. The authors report promising results, demonstrating the effectiveness of the proposed approach in the classification of EEG signals.

These work demonstrate the potential of EEG-based BMIs for mental and emotional sentiment classification. Both the machine learning-based approach and the deep evolutionary approach have shown promise in this area, enabling the development of more intelligent and reliable systems that can interpret and respond to human mental states in real-time. However, challenges remain, particularly in the selection of features and optimization of classifiers, which can significantly impact the accuracy and reliability of the system.

3 Methodology

3.1 Dataset

The data is collected from two people (1 male, 1 female) for 3 minutes per state (positive, neutral, negative). A Muse EEG headband is used, which recorded the TP9, AF7, AF8 and TP10 EEG placements via dry electrodes. The experimental protocol involved recording six minutes of neutral resting data, while inducing emotional states of positive and negative valence through the presentation of film clips for one minute per session. Additionally, a minute of neutral resting data was recorded without any stimuli. To obtain a comprehensive dataset, statistical analysis was performed to extract alpha, beta, theta, delta, and gamma brainwaves, resulting in a total of 2548 features.The dataset is freely available here. The figure below shows the distribution of data into various categories.

Fig. 1: Distribution of all 3 categories of the entire dataset

3.2 Preprocessing

Muse LSL is a Python package for streaming, visualizing, and recording EEG data from the Muse devices developed by InteraXon. This package is used to convert the data into csv file. This csv file is then resampled to generated a dataset by running over a script which is taken from `https://github.com/jordan-bird/eeg-feature-generation`. The final data consist of 2132 rows with 2548 features. The class distribution of the dataset was even i.e. nearly 700 records for each class.

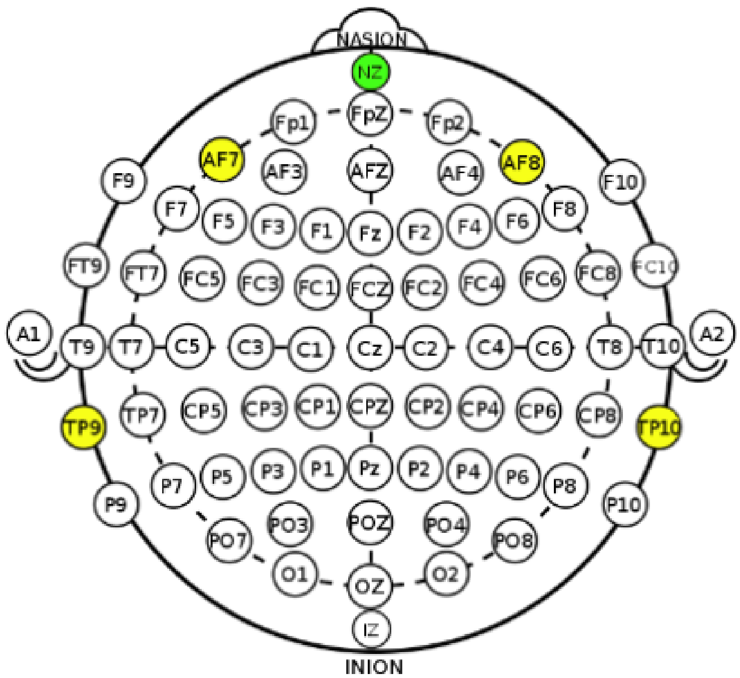


Fig. 2: Source of Film Clips used as Stimuli for EEG Brainwave Data Collection

Stimulus	**Valence**	**Studio**	**Year**
Marley and Me	Neg	Twentieth Century Fox, etc.	2008
Up	Neg	Walt Disney Pictures, etc.	2009
My Girl	Neg	Imagine Entertainment, etc.	1991
La La Land	Pos	Summit Entertainment, etc.	2016
Slow Life	Pos	BioQuest Studios	2014
Funny Dogs	Pos	MashupZone	2015

Fig. 3: EEG sensors TP9, AF7, AF8 and TP10 of the Muse headband on the international standard EEG placement system

3.3 Machine Learning Algorithms

Machine learning models are widely used for classification in a variety of fields, such as computer vision, natural language processing, and bioinformatics. These models can be trained using a variety of algorithms, including decision trees, random forests, support vector machines, and neural networks, among others. The choice of algorithm depends on the specific problem, the size and complexity of the data, and the performance metrics of interest. With advances in machine learning and access to large datasets, the accuracy and performance of classification models continue to improve, making them an essential tool for solving many real-world problems.
In this paper we have explored many such algorithms with multiple variations in hyper-parameter to optimize the performance on the dataset. We performed Principal Component Analyses (PCA) (a dimentionality reduction method) to study the feature importance with the variance within the data. Similarly the below methods were performed for classification:

Principle Component Analysis : Principal Component Analysis (PCA) is a dimensionality reduction technique that transforms high-dimensional data into a lower-dimensional space. PCA identifies the directions of maximum variance in the data and projects the data onto a new coordinate system defined by these directions, reducing the number of features while retaining as much of the original information as possible.

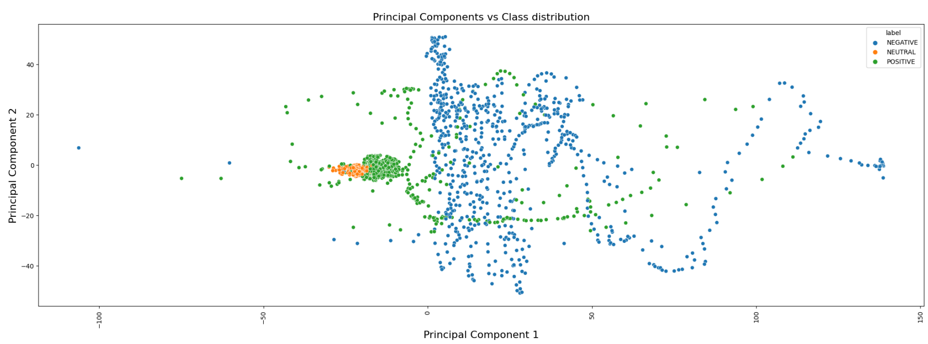


Fig. 4: Principal Component vs Class Distribution

Logistic Regression : Logistic regression is a statistical method commonly used for binary classification problems where the output variable has only two possible outcomes.It can be also extended to multiclass classification. It models the probability of the output variable using a sigmoidal function and can be trained using maximum likelihood estimation or gradient descent algorithms[7]. The logistic function is of the form: $p(x) = \dfrac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$ where $\beta_0 = -\mu/s$ and is known as the intercept (it is the vertical intercept or y-intercept of the line $y = \beta_0 + \beta_1 x$), and $\beta_1 = 1/s$ these are the y-intercept and slope of the log-odds as a function of x. Using scikit-learn library we normalize the data using standardscaler library and created a pipeline for logistic regression where various number of PCA component were used with max iteration size of 200. The cross validation score is illustrated in the table below:

Table 1: Performance of Logistic Regression

Number of PCA component	Cross-Validation Score
2	0.77
10	0.86
20	0.89

Random Forest : Random Forest is a powerful algorithm for multi-class classification tasks, where the goal is to classify instances into one of three or more classes[8]. To perform multi-class classification, Random Forest extends the binary decision tree model to handle multiple classes by constructing multiple trees and aggregating their predictions using a voting scheme. The algorithm is efficient, scalable, and can handle high-dimensional data, making it well-suited for many real-world multi-class classification problems. Here also we used variation in PCA component size to compute the score.

Table 2: Performance of Random Forest

Number of PCA component	Max Depth	Cross-Validation Score
2	5	0.92
10	5	0.92
20	5	0.93

Long Short-Term Memory (LSTM) : LSTM can be used for the classification of EEG data by modeling the temporal dependencies in the EEG signals and extracting relevant features from them[5]. By training an LSTM network on a large dataset of labeled EEG signals, it can learn to classify different EEG patterns, such as different types of brainwaves, and make accurate predictions on new, unlabeled EEG signals.

The entire data was split into 2 parts where 70% of the data was used for train purpose and remaining 30% data was used for testing the performance. The 70% of test data was again split into 70% train and 30% validation dataset. The labels in all these dataset component was one-hot encoded and saves as variable y variable for test , train and validate. A LSTM with 256 memory unit was used for this classification. The top layer consist of softmax activation function with output dimension of 3. This helps to calculate the probability of the data being among all the 3 class of positive, negative and neutral. The total number of trainable parameters in this model was 2,221,059. Adam was used as optimizer for training this lstm model with a learning rate of 0.001 but this learning rate was made dynamic based on the convergence of the model. Also, the model was trained for 50 epochs.

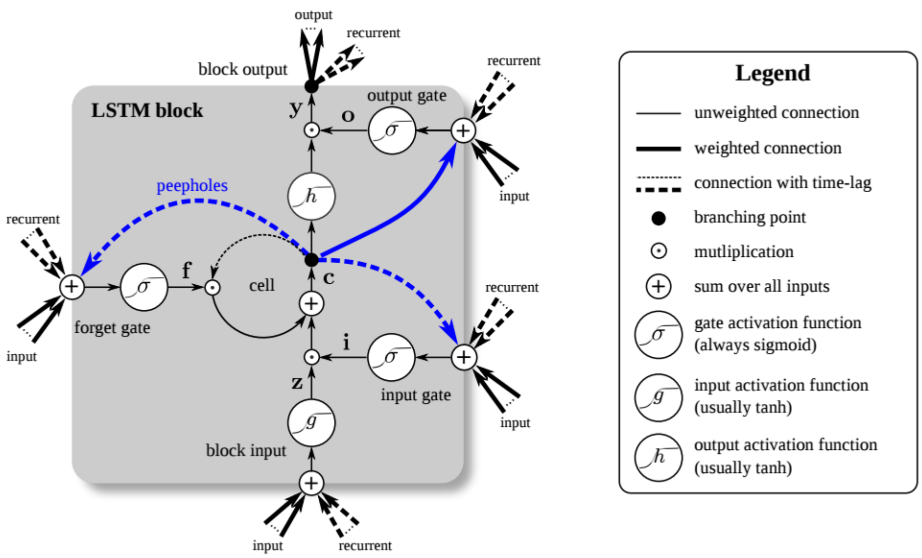


Fig. 5: Diagram of a standard block within a Long Short-Term Memory network

Gated Recurrent Unit (GRU) : GRU is a type of Recurrent Neural Network (RNN) architecture that can be used for classification of EEG data. GRU can model the temporal dependencies in EEG signals, and can learn to extract relevant features from them for classification of various EEG patterns, such as different types of brainwaves or emotional states. The data train, test and validation splits are same as the splits on lstm model and similar one-hot encoding is done for the target label. The model consists of three layers: an input layer, a GRU (Gated Recurrent Unit) layer, and a dense layer. The input layer expects input of shape (None, 2548, 1), meaning it can accept an arbitrary number of samples (denoted by None) with 2548 time steps and 1 feature dimension. The GRU layer has 256 units and is set to return sequences. It has 198,912 trainable parameters. The output of the GRU layer is flattened using the Flatten layer, resulting in a shape of (None, 652,288). Finally, the output is passed to a dense layer with 3 output units, which is used for classification. This layer has 1,956,867 trainable parameters. The

total number of trainable parameters in the model is 2,155,779. The model is optimized using a loss function during training to minimize the difference between predicted and actual output.
The Adam optimizer is used with a learning rate of 0.001, and the model is compiled with categorical cross-entropy loss and accuracy metric. The model is trained for 50 epochs with a batch size of 32. Early stopping and model checkpoint callbacks are used to prevent overfitting and save the best performing model, respectively. Additionally, a learning rate scheduler is implemented to adjust the learning rate during training. The training history is stored in the variable gruh, which contains the loss and accuracy values for each epoch.

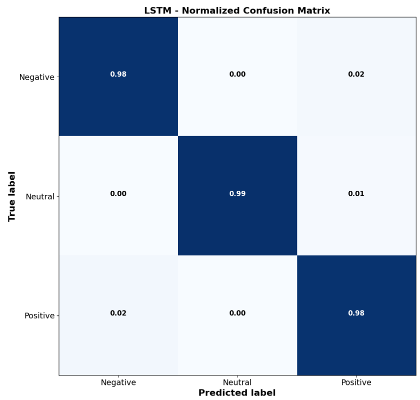
4 Results

Multiple suitable models were explored for emotion classification, including both machine learning and deep learning models. The machine learning models were pipelined with reduced dimension features using PCA, while the deep learning models were allowed to learn the features themselves. The best performing model was found to be the LSTM model with an accuracy of 98%. The EEG data with sliding window is a time-series data where the output is dependent on the previous state, therefore, the LSTM model was particularly suitable for this task due to its memory cell structure. The forget gate of the LSTM model allowed it to store important information and forget unnecessary information, resulting in high accuracy in this task.
Although the ensemble method, Random Forest, with reduced dimensionality demonstrated a high accuracy score, logistic regression failed to achieve the expected score. This could be attributed to the dataset's high dimensionality and complexity. Additionally, noise can often be present at the start of data readings.
The graph in fig 6 shows the accuracy and loss for LSTM model.

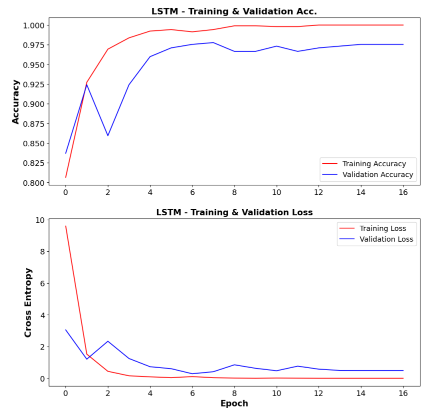
Table 3: Evaluation on test data for LSTM

CLASS	precision	recall	F1 Score	Support
0	0.98	0.98	0.98	216
1	1.00	0.99	1.00	215
2	0.97	0.98	0.97	209

Similarly, as compared to traditional recurrent neural networks (RNNs), **GRUs** have fewer parameters and are less prone to overfitting[3], which is especially important for emotion classification tasks where the amount of labeled data is often limited. Furthermore, GRUs have been shown to outperform other recurrent neural network architectures, such as LSTMs, on some time-series



(a) Confusion Matrix on LSTM model



(b) Graph of Accuracy and loss of LSTM model

Fig. 6: Performance on LSTM model

classification tasks, including emotion classification from EEG signals. In this experiment the GRU has resulted in accuracy of 96%. The detailed analysis is reported in table 4.

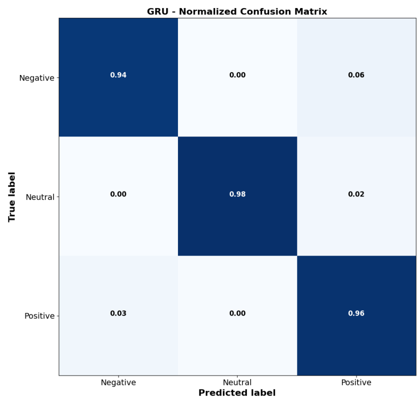
Table 4: Evaluation on test data for GRU

CLASS	precision	recall	F1 Score	Support
0	0.97	0.94	0.96	216
1	1.00	0.98	0.99	215
2	0.92	0.96	0.94	209

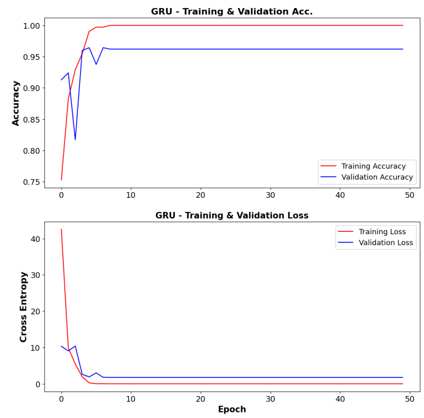
The Confusion matrix and graph of accuracy and loss is illustrated in fig 7(a) and 7(b) respectively.

5 Discussion

The success of deep learning models indicates that even more intricate models may provide significant benefits. Convolutional Neural Networks (CNNs) have demonstrated excellent performance in various classification tasks where data is sequentially or relatively connected. These techniques have the potential to deliver outstanding results. Additionally, ensemble and Bayesian models are promising directions that could produce better results when used in combination with advanced models, such as Dynamic Bayesian Mixture Models (DBMM),



(a) Confusion Matrix on GRU model



(b) Graph of Accuracy and loss of GRU model

Fig. 7: Performance on GRU model

which have previously been used to analyze statistical data derived from EEG brainwave signals.

The ability to automatically recognize emotions could prove beneficial for mental health decision support systems like GRiST, which is a safety management system utilized by mental health practitioners and individuals for self-assessment[4]. Being able to assess emotions without relying on self-reporting would enhance the accuracy of the advice and guide more sensitive interactions. Also , it can be used in the field of education to study the emotion of student during class hours and the changes when dealing with complex topics in the class.

References

[1] Jordan Bird et al. “A Deep Evolutionary Approach to Bioinspired Classifier Optimisation for Brain-Machine Interaction”. In: Aug. 2019.

[2] Jordan Bird et al. “Mental Emotional Sentiment Classification with an EEG-based Brain-machine Interface”. In: Apr. 2019.

[3] Kyunghyun Cho et al. *Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation*. 2014. arXiv: `1406.1078 [cs.CL]`.

[4] Diego R. Faria et al. “Probabilistic Human Daily Activity Recognition towards Robot-Assisted Living”. In: *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. Kobe, Japan: IEEE Press, 2015, pp. 582–587. DOI: `10.1109/ROMAN.2015.7333644`. URL: `https://doi.org/10.1109/ROMAN.2015.7333644`.

[5] Alex Graves, Navdeep Jaitly, and Abdel-rahman Mohamed. "Hybrid speech recognition with Deep Bidirectional LSTM". In: *2013 IEEE Workshop on Automatic Speech Recognition and Understanding*. 2013, pp. 273–278. DOI: 10.1109/ASRU.2013.6707742.

[6] Ruchika Malhotra, Dhananjay Deswal, and Jai Chaudhry. "Decoding the Brain Waves Using EEG Signals for Classifying Body Gestures by Applying Suitable ML amp; DL Techniques". In: *2022 4th Asia Pacific Information Technology Conference*. APIT 2022. Virtual Event, Thailand: Association for Computing Machinery, 2022, pp. 116–122. ISBN: 9781450395571. DOI: 10.1145/3512353.3512370. URL: https://doi.org/10.1145/3512353.3512370.

[7] Joanne Peng, Kuk Lee, and Gary Ingersoll. "An Introduction to Logistic Regression Analysis and Reporting". In: *Journal of Educational Research - J EDUC RES* 96 (Sept. 2002), pp. 3–14. DOI: 10.1080/00220670209598786.

[8] Chathurika S Wickramasinghe et al. "Generalization of deep learning for cyber-physical system security: A survey". In: *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*. IEEE. 2018, pp. 745–751.