UNIVERSITY OF PIRAEUS

School of Information and Communication Technologies

Department of Informatics

Thesis Title	**Evaluating the Efficacy of Machine Learning Models in Emotion Recognition through EEG Data Analysis**
Τίτλος Πτυχιακής Εργασίας	**Αξιολόγηση της αποτελεσματικότητας των μοντέλων μηχανικής μάθησης στην αναγνώριση συναισθημάτων μέσω ανάλυσης δεδομένων EEG**
Student's name-surname	**Emmanouil Moraitis**
Father's name	**Stamatios**
Student's ID No	**Π19275**
Supervisor	**Dionisios Sotiropoulos**

January 2025

The opinions and conclusions contained in this document are solely those of the author and do not represent the official positions of the University of Piraeus.

As the author of this paper, I declare that this paper is not a product of plagiarism and does not contain material from unmentioned sources.

1 Acknowledgments

First and foremost, I would like to express my heartfelt gratitude to my advisor, Dionisios Sotiropoulos, for their guidance, encouragement, and support throughout the journey of this thesis. Your expertise, patience, and constructive feedback have been very important in shaping this thesis and understanding of the subject.

I am also deeply thankful to my family, whose love, support, and belief in me have been a constant source of strength. To my mother, aunt and sister, thank you for always being my pillars of encouragement and for instilling in me the value of perseverance and hard work.

This thesis reflects the collective support and encouragement I have received from my teacher and my family, and I am profoundly grateful for their role in helping me reach this milestone.

2 Table of Contents

3 Table Of Figures

4 Abstract

Emotion recognition using EEG data is a field that stand between computing and neuroscience. This study evaluates the performance of four machine learning models, Support Vector Machine (SVM), Multi-Layer Perceptron (MLP) Classifier, Logistic Regression, and Decision Tree Classifier classifying emotional states (NEGATIVE, POSITIVE, NEUTRAL) based on EEG signals. Leveraging robust preprocessing techniques and K-Fold cross-validation, we assessed each model's efficacy in handling the complexity and high dimensionality of EEG data.

The results proved that SVM and MLP Classifier outperformed others with high accuracy, with very little misclassification in all classes, while Logistic Regression and Decision Tree Classifier, although interpretable, showed limitations in modelling subtle emotional patterns and, especially, in class differentiation for overlapping classes. The identified challenges are noise in EEG signals, model overfitting, and the trade-off between model performance vs interpretability, which have been addressed to provide insight into model suitability for real-world applications.

This study further develops emotion recognition technologies, establishing a baseline for machine learning approaches applied to EEG data. It also opens paths to further investigation.

By combining multiple models, fine-tuning the features they use, and taking changes in emotions over time into account, we can make emotion recognition systems more dependable and easier to apply. This will be especially helpful in fields like mental health, adaptive interactions between people and computers, and personalized learning programs.

Key Words: Electroencephalography (EEG), Emotion Recognition, Machine Learning, Neural Networks, Data Classification.

5 Περίληψη

Η αναγνώριση συναισθημάτων μέσω δεδομένων EEG είναι ένα πεδίο που βρίσκεται στο μεταίχμιο της πληροφορικής και της νευροεπιστήμης. Η παρούσα μελέτη αξιολογεί την απόδοση τεσσάρων μοντέλων μηχανικής μάθησης, Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), Logistic Regression και Decision Tree Classifier, στην ταξινόμηση συναισθηματικών καταστάσεων (ΑΡΝΗΤΙΚΟ, ΘΕΤΙΚΟ, ΟΥΔΕΤΕΡΟ) βάσει σημάτων EEG. Χρησιμοποιώντας ισχυρές τεχνικές προεπεξεργασίας και διασταυρούμενη επικύρωση K-Fold, εκτιμήθηκε η ικανότητα κάθε μοντέλου να διαχειριστεί την πολυπλοκότητα και την υψηλή διαστασιμότητα των δεδομένων EEG.

Τα αποτελέσματα έδειξαν ότι τα μοντέλα SVM και MLP ξεπέρασαν σε απόδοση τα υπόλοιπα, επιτυγχάνοντας υψηλή ακρίβεια και ελάχιστη λανθασμένη ταξινόμηση σε όλες τις κατηγορίες. Αντίθετα, Logistic Regression και Decision Tree Classifier, παρότι πιο ερμηνεύσιμα, παρουσίασαν περιορισμούς στη μοντελοποίηση λεπτών συναισθηματικών μοτίβων και, ιδιαιτέρως, στη διαχωριστική ικανότητα μεταξύ επικαλυπτόμενων κατηγοριών. Οι προσδιορισμένες προκλήσεις περιλαμβάνουν τον θόρυβο στα σήματα EEG, την υπερπροσαρμογή των μοντέλων και το δίλημμα μεταξύ απόδοσης και δυνατότητα ερμηνείας, τα οποία αντιμετωπίστηκαν, παρέχοντας χρήσιμες πληροφορίες για την καταλληλόλητα των μοντέλων σε πραγματικές εφαρμογές.

Η παρούσα μελέτη προάγει περαιτέρω τις τεχνολογίες αναγνώρισης συναισθημάτων, θέτοντας ένα σημείο αναφοράς για προσεγγίσεις μηχανικής μάθησης που εφαρμόζονται σε δεδομένα EEG, και ταυτόχρονα ανοίγει τον δρόμο για περαιτέρω έρευνα.

Με τον συνδυασμό πολλαπλών μοντέλων, τη βελτιστοποίηση των χαρακτηριστικών που χρησιμοποιούν και την ενσωμάτωση της χρονικής μεταβολής των συναισθημάτων, μπορούμε να κάνουμε τα συστήματα αναγνώρισης συναισθημάτων πιο αξιόπιστα και ευκολότερα στην εφαρμογή. Αυτό θα αποδειχθεί ιδιαίτερα χρήσιμο σε τομείς όπως η ψυχική υγεία, οι προσαρμοστικές αλληλεπιδράσεις μεταξύ ανθρώπων και υπολογιστών, καθώς και τα εξατομικευμένα εκπαιδευτικά προγράμματα.

Λέξεις Κλειδιά : Ηλεκτροεγκεφαλογράφημα (EEG), Αναγνώριση Συναισθημάτων, Αναγνώριση Συναισθημάτων, Νευρωνικά Δίκτυα, Ταξινόμηση Δεδομένων.

6 Introduction

6.1 What is Machine Learning?

Machine learning is a branch of artificial intelligence (AI) that focuses on teaching computers how to learn and make decisions from data. Instead of being explicitly programmed to perform a task, a machine learning model learns patterns and relationships in the data it is given, enabling it to make predictions or decisions on new, unseen data [4], [21]. Think of it as training a computer the way you might train a pet, by showing it examples and rewarding it for getting things right.

6.1.1 Why is Machine Learning Important?

Machine learning is transforming industries and research fields because it can process and analyze vast amounts of data more efficiently than humans [21], [5]. In emotion recognition using EEG (electroencephalographic) data, machine learning models help decode complex brainwave patterns, making it possible to classify emotions with high accuracy [13], [7]. This has potential applications in mental health, brain-computer interfaces, adaptive technologies, and more.

Machine learning is all around us. From the recommendations on Netflix to the voice recognition on your phone, its application to understanding emotions is just one of many ways this technology is shaping the future.

6.2 Historical Background of Emotion Recognition

The exploration of emotion recognition through EEG (electroencephalographic) data is rooted in the intersection of neuroscience, psychology, and computing. Historically, the study of emotions began with philosophical and psychological inquiries into the nature of human feelings and their influence on behavior. Scientists like William James and Carl Lange in the late 19th century laid the groundwork by proposing that emotions arise from physiological changes in the body, paving the way for research into the biological underpinnings of emotions [13], [15].

In the 20th century, advancements in neuroscience allowed for a deeper understanding of the brain’s role in emotions. The discovery of the limbic system as a central region involved in emotional processing was a milestone [15]. During this time, EEG technology emerged as a tool to measure electrical activity in the brain. Initially, EEG was used for medical purposes, such as diagnosing epilepsy and sleep disorders [12]. However, researchers soon recognized its potential for studying cognitive and emotional processes [13], [7].

The first attempts to use EEG for emotion recognition began in the mid-20th century when scientists observed changes in brainwave patterns associated with emotional states. Early studies focused on identifying correlations between EEG frequency bands (such as alpha, beta, theta, and gamma waves) and basic emotional responses like happiness, fear, or anger [13], [18]. These efforts were largely constrained by limited computational resources and the rudimentary nature of EEG recording devices.

The late 20th century saw the rise of computational tools, which significantly advanced the field. Researchers began applying statistical methods and basic machine learning algorithms to classify EEG signals. For example, Support Vector Machines (SVMs) and k-Nearest Neighbors (k-NN) were employed to distinguish between emotional states based on features extracted from EEG data [5], [7].

In the 21st century, the convergence of neuroscience, computing, and artificial intelligence revolutionized emotion recognition using EEG. The introduction of high-resolution EEG devices and open-access EEG datasets provided researchers with the resources to conduct large-scale studies [1], [9]. At the same time, advancements in machine learning, particularly deep learning, allowed for the modeling of complex and non-linear patterns in EEG signals. Algorithms like Feedforward Neural Networks (FFNNs) and Convolutional Neural Networks (CNNs) began outperforming traditional methods, achieving higher accuracy in recognizing emotions [13], [14].

Today, EEG-based emotion recognition is a rapidly evolving field. It is being applied in mental health monitoring, human-computer interaction, and adaptive AI systems [16], [18]. Despite significant progress, challenges such as noise in EEG signals, variability across individuals, and ethical considerations remain [6], [17]. These issues continue to drive research into more robust and equitable solutions, ensuring that the historical journey of EEG-based emotion recognition is far from over.

6.3 The Significance of Emotion Recognition Using EEG Data

Emotion recognition from EEG has great potential in the field of monitoring and assessment related to mental health. EEG records brain activity in real time, which enables one to display superficial pointers of emotions, indicated by changes in facial expression or the timbre of one's voice, with deeper, more accurate reflections of emotional states [13],[16]. This is real-time insight into tool development that might avail real-time support within a mental health context, hence permitting more precise and responsive intervention for persons under stress, anxieties, or any emotional setbacks [16],[17].

This research is also driven by a vision for developing AI systems similar in nature to the operation mode of a human brain [4],[21]. With this, we are teaching AI to identify all the different types of emotions based on their brainwave patterns, thus inching a bit closer to making it understand and intuitively interact with humans [7],[14]. That is where AI in the mainstream becomes not just functionally intelligent but sympathetic or responsive to the human experience [6]. Emotion recognition through EEG is just such an important bridge between neuroscience and AI, offering progress which might reshape the way machines will understand and respond to human emotions.

6.4 The Potential of SVMs, FFNN, Logistic Regression, and Decision Trees in EEG Analysis

SVMs, FFNNs, logistic regression, and decision trees are some of the most important contemporary machine learning models used in the analysis of electroencephalographic data. Each model has advantages in interpreting the complex patterns inherent in EEG signals.

A support vector machine can classify EEG data by finding the most efficient hyperplane that classifies different classes. Therefore, the capability to handle high-dimensional data makes it appropriate for distinguishing various states of a subject's mind [7],[13]. For example, in epileptic seizure detection from EEG signals, it has shown very strong performance in classification [12].

Feed-forward neural networks (FFNNs) are good at modeling complex relations within EEG data. The architecture enables them to learn even intricate patterns, especially those that are important for tasks like emotion recognition or cognitive state monitoring [14],[13]. There is evidence from various research showing that FFNNs can classify various EEG signals related to different mental tasks, which may help in the development of useful brain-computer interfaces [7],[16].

Logistic regression is one of the simplest but boundlessly used statistical models in EEG data analysis. For instance, in binary classification, such as the classification of EEG signals for normal and abnormal brain activities, the model has been quite effective [5],[9]. Feature selection and classification using logistic regression provide high performance for EEG-based motor imagery tasks identified in the literature [5],[18].

Decision trees represent an understandable and interpretable methodology for the classification of EEG data. A decision tree operates by recursively splitting the data depending on the value of the features, producing a tree-like model for decisions [10]. This is applied in the detection of epileptic seizures and gives clear insight into the process [12],[18].

In summary, some of the very important and critical machine learning models in EEG analysis are discussed here, each performing unique functions in the analysis and classification of brain signals. The application of their uses supports high-value advanced diagnosis and therapeutic tool development in the understanding of neural activities.

7 Literature Review

7.1 Overview of EEG Data and Emotion Recognition

EEG is a non-invasive process in which the electrical activity of the brain is measured with electrodes attached to the scalp. It detects voltage changes caused by ionic current flows within neurons, reflecting real-time activity in the brain [15],[12]. The EEG signal consists of multiple frequency bands. Each frequency band is associated with different states of cognition and emotion [13],[18].

Emotion recognition using EEG consists of the analysis of such brainwave patterns to identify and classify human emotions. It includes the following general steps:

- **Data Acquisition:** Gathering EEG signals while subjects are exposed to various emotional stimuli. Preprocessing steps include filtering and cleaning the raw EEG to remove noise and artifacts [1],[13].
- **Feature Extraction:** To identify relevant features from the EEG signals, which are related to an emotional state [7],[18].
- **Classification:** Application of machine learning algorithms for categorizing the extracted features under specific emotions [5],[7].

The contributions of machine learning techniques like SVM, FFNN, logistic regression, and decision trees have gone a long way in developing accuracy for emotion recognition systems [5],[7],[13]. It is now possible to learn complicated patterns of EEG data with better models, thus enabling the better identification of emotional states [14],[18].

As a result, these improvements give rise to real-world applications in fields such as mental health monitoring, human-computer interaction, and adaptive learning systems [16],[18]. An example of this is how EEG-based emotion recognition may form the basis for emotional disorder diagnosis or personalized experience regulation in real-time [13],[16].

However, challenges persist in the form of requiring large, varied datasets or the development of models that generalize well across a wide range of individuals. Active research studies focus on overcoming these vital challenges in developing more robust and reliable emotion recognition systems [13],[6].

8 Justification for Model Selection

The models' choice for selection in the present research was guided by their appropriateness for analyzing EEG data and their proven performance in tasks related to emotion recognition.

Support Vector Machines (SVMs) were chosen because of their capability to handle high-dimensional feature datasets, which is very important given the complexity of the nature of EEG signals [7],[13]. Their strength in estimating hyperplanes for class separation makes them particularly competent for tasks of complex classification, whereby clear distinctions between different emotional classes need to be made [18].

Feedforward Neural Networks (FFNNs) were included due to their capacity to learn complex, nonlinear relationships that might exist within the data. Such a property is highly useful in emotion recognition, as the underlying patterns within the EEG data may be very subtle and require higher-order analysis methodologies to extract meaningful information. FFNNs are very good at unearthing such complex dynamics; therefore, they form an integral part of the analytical pipeline in this work [13],[14].

While intuitive in its simplicity, **Logistic Regression** was preferred as it offers interpretability for both binary and multi-class classification problems [5],[9]. This simplicity does not limit its effectiveness but proves quite helpful for understanding the linear separability inherent in the data and providing a fundamental baseline against which other complex models are compared [5].

Decision Trees were opted for because they are straightforward, easy to understand, and allow clear visualization of decisions based on the features of EEG [10],[18]. Their application in handling feature-based classification tasks adds significant value in terms of information about the importance and contribution of each feature toward changes within the dataset [10],[12].

The integration of these models ensures a holistic approach in analyzing emotion recognition based on EEG data. Each model has its merits, from the interpretive strength of Logistic Regression and Decision Trees to the advanced pattern recognition capabilities of Support Vector Machines and Feedforward Neural Networks. This combination enables an exhaustive evaluation of the research objectives.

8.1 Applications of SVMs, FFNN, Logistic Regression, and Decision Trees in EEG Data Analysis

Several machine learning algorithms have been in wide application for EEG data analyses, including SVMs, FFNNs, logistic regression, and decision trees, each presenting advantages in deciphering these complex brain signals.

SVMs can perform the classification of EEG data by finding an optimal hyperplane separating different classes [7],[18]. Since they can handle high-dimensional data very well, they are used in distinguishing between different mental states [13]. For instance, SVM has been used to detect seizures from an EEG signal, so it is appropriate and very effective for classification [12].

FFNNs, an important class of artificial neural networks, have strong capabilities in modeling complex relationships within EEG data. Their layered architecture allows them to learn intricate patterns, turning them useful in tasks related to emotion recognition and the monitoring of one's cognitive state [13]. It has been proven that FFNNs classify EEG signals corresponding to different mental tasks efficiently, thus contributing to the development of brain-computer interface technologies [14].

The Logistic Regression model is simple and has also been adopted in the analysis of EEG data [5],[9]. This model is applied in areas such as binary classification problems, like discrimination between normal and abnormal brain activity [9]. It is used in the processes of feature selection and classification of EEG recordings for subjects performing motor imagery tasks, demonstrating very good performance [18].

Decision Trees provide a transparent and interpretable approach to EEG data classification. They recursively split the data based on feature values, yielding a tree-like model of decisions [10]. This approach has been used for epileptic seizure detection and yields good insight into the decision-making process [12],[18].

These machine learning models hence form the backbone of EEG analysis, as each model contributes in a different manner to the interpretation and classification of brain signals. Their application enhances learning about neural activities and developing further diagnostics and therapeutic tools.

9 Opportunities and Challenges in Emotion Recognition Using EEG Data

The study of emotion recognition from EEG data is an area rich with potential, both in terms of expanding our understanding of cognition and emotion in humans as well as developing new innovations in healthcare and human computer interaction but also artificial intelligence. Yet, it is also grappling with major issues that must be resolved to unlock its potential.

9.1.1 Opportunities

Improved human-machine interaction:

Emotion recognition within HCI (Human-Computer Interaction) could make systems more intuitive and responsive [16],[18]. For example, adaptive learning platforms could adapt content delivery according to the emotional state of the learner, yielding better learning engagement and retention [13],[14].

By using EEG to recognize emotions without any invasive procedures, it becomes much easier and stress-free to track a person's emotional state [7],[13]. This can help doctors and caregivers spot issues like depression and anxiety earlier, leading to better treatment and support [6],[17].

Affective Computing Advances:

These emerging technologies are steadily being integrated into systems that not only detect and interpret human emotions but also respond to them in meaningful ways [13],[16]. By incorporating a deeper understanding of users' emotional states, these systems can greatly enhance user experiences across a broad spectrum of applications—ranging from mental health support and adaptive learning tools to personalized entertainment and customer service [16],[18]. In doing so, they introduce a powerful dimension to human-computer interaction, making technology more empathetic, engaging, and ultimately more supportive of human well-being.

9.1.2 Challenges

EEG signals are unique to the individual and may change under many conditions, such as with age, gender, and environmental conditions. This makes it hard to create models that are more generalizable for any population.

Signal Noise and Artifact:

Muscle movements, eye blinks, and other forms of environmental electrical sources are usually a kind of noise in EEG data that masks the neural signals corresponding to emotions [12],[13]. There is some complexity of emotional states, usually depicted by heavily overlapping neural patterns, that make the classification of distinct emotional states from EEG data complicated [7],[18]. These issues call for further research in the enhancement of data preprocessing methods, creation of robust machine learning algorithms, and establishment of standard mechanisms regarding the elicitation of emotions and acquisition of EEG data [13],[14]. Only through these advancements can the full potential of EEG-based emotion recognition be realized and applied.

10 Methodology

10.1 Justification for the Selection of SVMs, FFNN, Logistic Regression, and Decision Trees

In the case of EEG-based emotion recognition, the selection of appropriate models is vital for capturing and interpreting the intrinsic patterns in brainwaves. Out of the various machine-learning model selections, including Support Vector Machines, FFNNs, Logistic Regression, and Decision Trees, each model has certain advantages that can guarantee its selection.

10.1.1 Support Vector Machine

Such high-dimensional data also tend to perform very well with the SVM and are very effective for binary classification problems [7],[18]. This algorithm locates the optimal separating hyperplane between classes, hence finding a lot of use in identifying different emotional states represented in EEG signals [13]. Kernel functions allow the application of SVMs in a wide circle of complex analyses of EEG records in case of non-linearity [14].

10.1.2 Feedforward Neural Networks

FFNNs are artificial neural network models that possess the ability to model complex nonlinear relationships within data [14],[13]. Their architecture lets them learn from and represent really intricate patterns within the EEG signal to recognize subtle variations in emotions. Because of the ease with which they can adapt for different input features, along with their deep learning capabilities, FFNNs serve as a very useful tool for EEG-based emotion recognition [7],[14].

10.1.3 Logistic Regression

While simple, Logistic Regression is a surprisingly powerful baseline classifier. It predicts the probability of a binary output given some input features and is highly interpretable, that is, necessary to understand which specific EEG features drive the emotional states [9]. Its efficiency and ease of implementation make it of practical use for the initial analyses and feature selection of EEG studies [18],[5].

10.1.4 Decision Trees

Decision Trees provide an extremely interpretable and understandable model for classification, such that the data are recursively partitioned based on feature values [10],[18]. This approach explains the reasoning behind decision-making; therefore, the researcher can specify which EEG feature is more relevant with respect to a specific emotion. Their easy-to-understand visualization allows the researcher to grasp the hierarchical relationships between features and their outcomes, which helps the researcher in the context of EEG data analysis [12]. These models together propose a complete framework for evaluation that capitalizes on their respective strengths in model development with both accuracy and interpretability for emotion recognition from EEG data [5],[13].

10.2 Description of K-fold Cross-Validation for Robust Assessment

K-fold cross-validation is considered one of the robust techniques for evaluating a machine learning model, especially when working with restricted datasets [11],[5]. It works by partitioning one dataset into K equally sized subsets, or "folds." For system training and validation in a model, K times run using a different fold every single time as the validation set with the rest K-1 folds as the training set [11]. It ensures that every data point gets used in both training and validation, thus allowing an overall assessment of the model performance [5],[9].

The process involved in K-fold cross-validation:

1. Partitioning of Data: Divide the dataset into K equal folds.

2. Model Training and Validation: Repeat for each fold:

- Use the current fold for the validation set.
- Take the remaining K-1 folds as the training set.
- Train the model on the training set.
- Use the model to make predictions on the validation set.

3. Performance Aggregation: Aggregate the performance metrics of the model, such as accuracy, precision, recall, over all K iterations to obtain an overall estimate of the model's effectiveness.

10.2.1 The advantages of this approach

Efficient Data Utilization: K-fold cross-validation assures that every data point has been used for training and validation; it maximizes the available data, which can be useful for small datasets [11],[21],[22].

Smaller Variability in Performance Estimates: Averaging the results over K iterations gives a more robust model performance estimate that is resistant to data variability [11],[21],[22].

Better Generalization: Measuring the model's performance on different validation sets provides more comprehensive insight into how well it generalizes to new, unseen data. This, in turn, reduces overfitting.

K-fold cross-validation holds special value in the context of EEG-based emotion recognition, where EEG data might be particularly complex and subject to variability because of individual differences and other conditions [13],[18]. Application of K-fold cross-validation allows for an extensive assessment of model performance on different subsets of data, hence allowing one to ensure that a model is robust and generalizes well on new, unseen EEG signals [11],[21].

11 Implementation

11.1 Code for Comparing Classification Accuracy: Implementation and Tool Usage

The complete thesis's code has been executed in Visual Studio Code (VSCode), which makes full utilization of its IDE capabilities, and the is written in python. Using important libraries such as:

- scikit-learn for machine learning [3],[8]
- NumPy for numerical operations [27],[28]
- pandas for data manipulation [25]
- seaborn, and matplotlib for visualization in the present EEG study [26],[29],[19]

Also, detailed k-fold cross-validation will be employed to thoroughly assess each model's appropriateness and generalization in classifying emotional states.

11.1.1 Preparing Data

The Python script, data_preparation.py, first reads the EEG dataset and performs some preliminary exploratory data analysis on the data.

This will be achieved by showing some of the initial rows in the data, describing the information in the datasets, giving descriptive statistics and verifying for any missing values.

The last step is standardizing features using a StandardScaler from scikit-learn [3],[8]. Standardization in the analysis of EEG data is crucial, this will ensure all features contribute equally toward training the model and preventing any single feature from dominating due to its scale.
This means that features with higher magnitude can "dominate" the learning process,

skewing the model's understanding and potentially leading to inaccurate predictions. Standardizing the features transforms the data so that each feature has a mean of zero and a standard deviation of one. This puts all features on the same scale, ensuring that each one contributes equally to the training of the model. It prevents any single feature from disproportionately influencing the model due to its numerical scale, allowing the model to learn more effectively from all the available data [5].

11.1.2 Model Initialization

The following machine learning models are initialized in model_config.py:

Support Vector Machine

Linear Kernel, C = 1.0. An SVM with this configuration is best indicated for data that is linearly separable.

FFNN

An MLPClassifier with one hidden layer consisting of 100 neurons, ReLU for the activation technique, and using the Adam optimizer was set up and would run for up to 300 iterations.

Logistic Regression

Uses a liblinear solver, L2 regularization, regularizing strength C=1.0 usually used for binary and multinomial classification problems with high efficiency.

Decision Tree

The maximum depth is 5, using the gini impurity criterion for balance between model generalization and overfitting.

K-Fold Cross-Validation

The key evaluation is done in train_test_models_k_fold where we execute K-fold cross-validation with n_splits=5, meaning that it will split the data into five subsets and use iteratively four of them for training and one for testing in such a way that each subset once becomes a test set.

That will bound the overfitting problem and yield ground for the realistic performance evaluation concerning different sets of data. In all cases, we present the per-fold accuracy and then the average because it is a better representative of the generalization power of a particular model.

Item-by-Item Review

We go beyond cross-validation with the classic 80/20 train-test split for further performance analysis of each model.

We call the function evaluate model which trains the model on a train set and then does a prediction on the test set.

Then it shows the confusion matrix and plot using seaborn heatmap to get an idea about the model capability for classification and to point out the regions of misclassification.

We also show a classification report with Precision, Recall, and F1-score for each of the classes, namely NEGATIVE, POSITIVE, and NEUTRAL, representing the overall performance of the model on different emotions.

12 Results

12.1 Visual Comparison of Model Performance

The following section shows, in a graphical manner, the performance comparison for four machine learning models, namely Support Vector Machine, Multi-Layer Perceptron Classifier, Logistic Regression, and Decision Tree Classifier, which classify the emotional states from EEG data. Each model was tested with the confusion matrix to show the correctness of distinguishing the NEGATIVE, POSITIVE, and NEUTRAL emotional states. These matrices show certain patterns in the classification accuracy of the models and also points where each may fail for a particular class, hence providing insight beyond mere performance metrics such as accuracy.

12.1.1 Confusion Matrix Analysis

The confusion matrix is an essential component in evaluating the model, as it provides a detailed breakdown of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) across classes. For every model, the confusion matrix graphically illustrates the instances which were classified correctly, the diagonal entries, and misclassifications, the off-diagonal entries. It thus gives us the chance to further assess the model's precision, recall, and F1-score for each emotion class, thereby giving more information about the strengths and weaknesses of the model [5],[18].

12.1.1.1 SVM (Support Vector Machine)

The SVM model demonstrates high accuracy, with minimal misclassifications across all classes. Its confusion matrix shows most instances correctly classified for NEGATIVE, POSITIVE, and NEUTRAL classes, indicating strong boundary definition in the feature space. This model has few errors, particularly between NEGATIVE and NEUTRAL classes, suggesting robust performance in distinguishing different emotional states.

```
Results for SVC:
Classification Report:
              precision    recall  f1-score   support

    NEGATIVE       0.97      0.98      0.97       143
    POSITIVE       0.99      0.99      0.99       148
     NEUTRAL       0.97      0.96      0.97       136

    accuracy                           0.98       427
   macro avg       0.98      0.98      0.98       427
weighted avg       0.98      0.98      0.98       427
```

Figure 1 Results for SVC

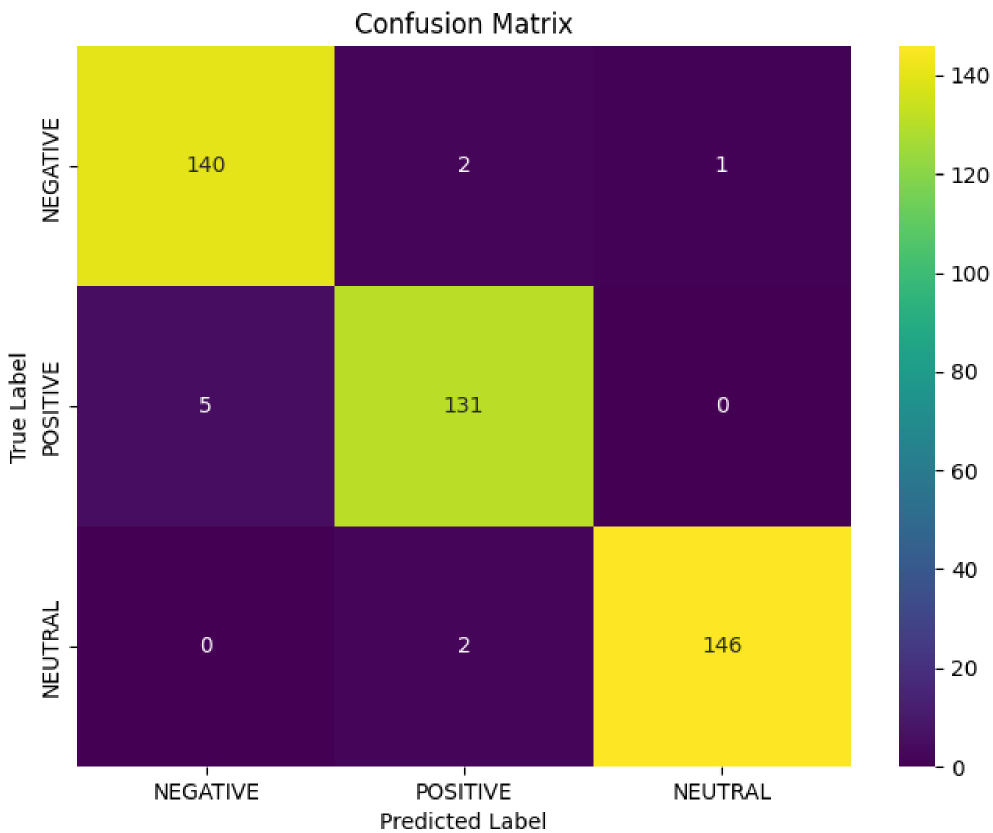

Figure 2 Confusion Matrix of SVC

12.1.1.2 MLP Classifier Feed-Forward Neural Network

Similarly, the MLP Classifier provides high precision and recall in each class. In its confusion matrix, it presents fewer misclassifications when compared to other models, such as a Decision Tree or Logistic Regression. It appears that nonlinear relationship modelling by the MLP is quite helpful in classifying each emotion correctly into its class, while the cross-class misclassifications are very slight.

```
Results for MLPClassifier:
Classification Report:
              precision    recall  f1-score   support

    NEGATIVE       0.97      0.99      0.98       143
    POSITIVE       1.00      0.99      0.99       148
     NEUTRAL       0.98      0.97      0.97       136

    accuracy                           0.98       427
   macro avg       0.98      0.98      0.98       427
weighted avg       0.98      0.98      0.98       427
```

Figure 3 Results for MLPClassifier

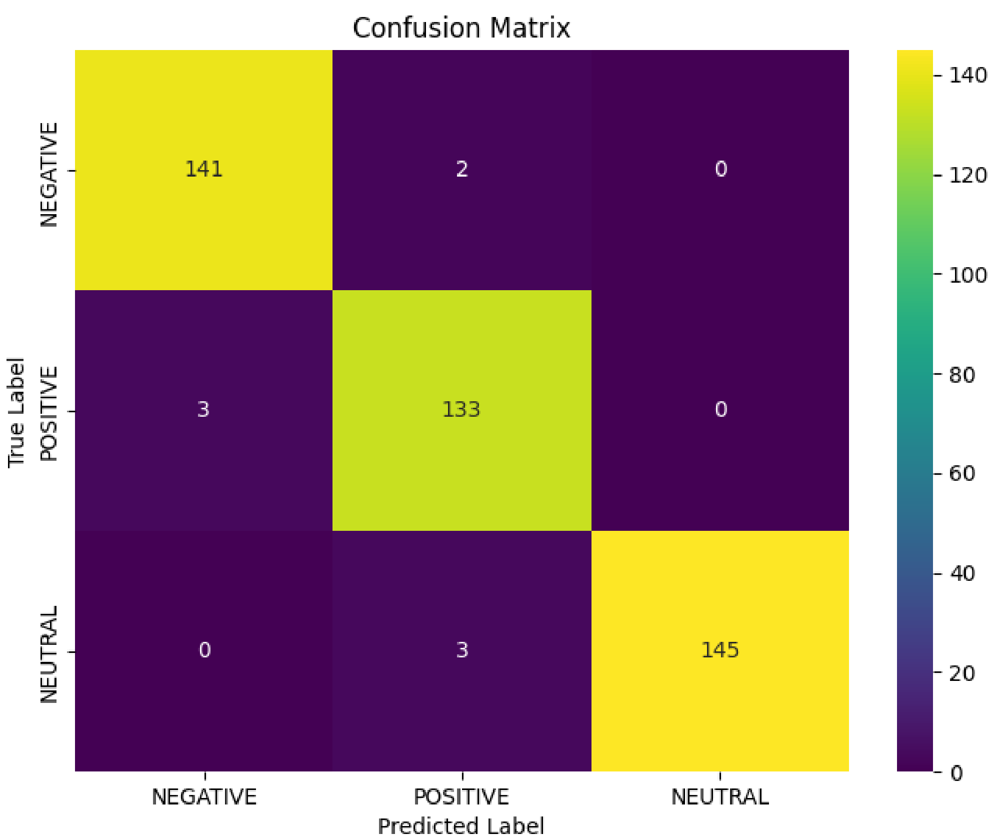

Figure 4 Confusion Matrix of MLPClassifier

12.1.1.3 Logistic Regression

Logistic Regression has behaved very well, particularly in the classification of the NEUTRAL class without any mistake. Its confusion matrix showed that minor misclassifications appeared in both NEGATIVE and POSITIVE classes. Although Logistic Regression has a linear decision boundary, which generally fits worse for complex patterns in EEG data, it keeps high overall accuracy. However, the limitation of this classifier appears when the subtle differences between the similar emotional states need to be differentiated, especially in NEGATIVE and POSITIVE categories.

Results for LogisticRegression:
Classification Report:

	precision	recall	f1-score	support
NEGATIVE	0.93	0.97	0.95	143
POSITIVE	1.00	1.00	1.00	148
NEUTRAL	0.97	0.93	0.95	136
accuracy			0.97	427
macro avg	0.97	0.97	0.97	427
weighted avg	0.97	0.97	0.97	427

Figure 5 Results for Logistic Regression

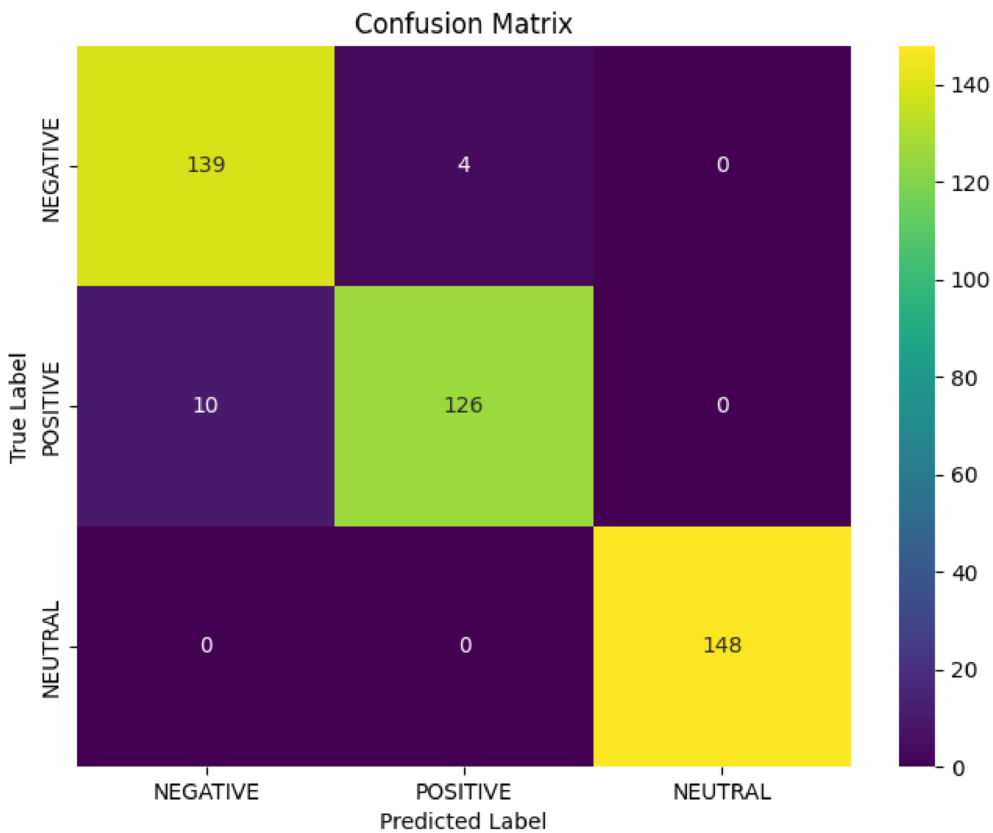

Figure 6 Confusion Matrix of MLPClassifier

12.1.1.4 Decision Tree Classifier

The Decision Tree model performs reasonably, although it tends to slightly higher misclassification rates compared to the other two models, namely the SVM and MLP models. The confusion matrix of this model points out occasional incorrect classification of POSITIVE instances as NEGATIVE and vice-versa, a probable consequence of overfitting to particular patterns in the training data used by decision trees. While easily interpretable, the simplicity of the Decision Tree may not capture, with the required complexity, the subtleties in the data of the EEG.

Results for DecisionTreeClassifier:
Classification Report:

	precision	recall	f1-score	support
NEGATIVE	0.93	0.97	0.95	143
POSITIVE	1.00	0.99	0.99	148
NEUTRAL	0.95	0.93	0.94	136
accuracy			0.96	427
macro avg	0.96	0.96	0.96	427
weighted avg	0.96	0.96	0.96	427

Figure 7 Results for Decision Tree Classifier

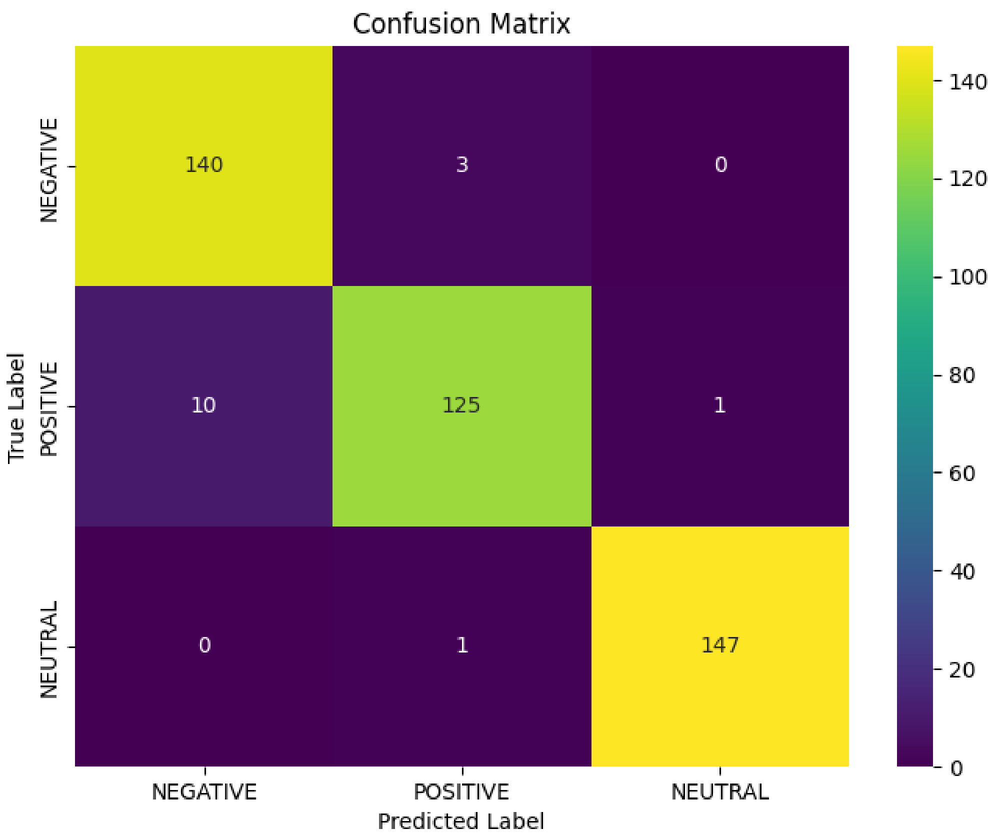

Figure 8 Confusion Matrix of Decision Tree Classifier

12.2 Interpretation of Results in Relation to Research Objectives

The main purpose of the research study is to compare machine learning models such as the Support Vector Machine, Multi-Layer Perceptron Classifier, Logistic Regression, and Decision Tree Classifier based on efficiency regarding the correct classification of emotional states from EEG data. Through the examination of quantitative performance for each model, it is possible to derive to what extent these models fulfil the goal of the given research-reliable emotion recognition based on neural signals.

12.2.1 Key Findings and Their Implications

12.2.1.1 Model Performance Overview

- **SVM and MLP** Classifier have emerged as the best models with very high accuracy, precision, and recall for every emotional state of NEGATIVE, POSITIVE, and NEUTRAL. These algorithms show very minimal misclassification of classes in their confusion matrices, especially for the NEUTRAL class. This would provide evidence that these models truly fit the capturing of complex patterns in EEG data. This will be important for applications where exact detection of emotions is at stake, such as monitoring mental health and adaptive human-computer interactions.

- **Logistic Regression** showed quite good performance and performed extremely well in the NEUTRAL class; it had no classification errors. However, the model sometimes got confused between NEGATIVE and POSITIVE states, probably due to the linear nature of Logistic Regression, which may fail to capture the subtle changes in EEG patterns manifesting different emotional states. This could therefore mean that although Logistic Regression is interpretable and its performance is relatively good, the model is a bit unsuitable for EEG emotion recognition as compared to nonlinear models such as SVM and MLP.

- On the other hand, **Decision Tree Classifier** presented an interpretable model, but it did show some misclassifications between NEGATIVE and POSITIVE emotions at a slightly higher rate. This might have been driven by overfitting, possibly due to the nature of the data itself, because EEG is usually noisy and variable in real-world applications.

12.2.1.2 Alignment with Research Objectives

The study sought to find those models that can classify EEG-based emotions correctly, with a focus on generalizability and least misclassification [5],[7]. These are best met by the objectives of SVM and MLP Classifier, since they give the best accuracy and balanced classification across all classes, indicating thereby that these models are capable of grasping the minute variations in EEG patterns that correspond to emotional states [13], [14]. Their robustness shows the possibility of these models for applications which need high reliability, like that of continuous emotion monitoring in mental health settings [16],[18].

While Logistic Regression and Decision Tree Classifier perform decently, their performance also underlines some limitation in handling the complexity of EEG data for the task of emotion recognition [9],[10]. The overall low recall and precision for some classes in these also underlines their inability, with respect to other models, to capture the fine relationships in EEG data. These findings again underline the importance of model selection in EEG analysis, since linear or simpler models may not perform optimally on tasks that require complex pattern recognition.

12.2.1.3 Impact of Model Selection on Practical Applications

Better performances exhibited by SVM and MLP models reflect the fact that nonlinear and complex classifiers can give better performance in EEG-based emotion recognition, since they have the capability to understand the intricate neural signatures related to various emotions [7],[13],[14]. This could prove useful for future research and practical applications when the reliability of emotion detection will be imperative. For example, in adaptive learning environments, such models could dynamically adjust content based on the state of the learner to improve engagement and learning outcomes [13],[16].

On the contrary, the results indicate that Logistic Regression and Decision Tree provide only preliminary or exploratory analyses in which understanding the underlying processes is more important than the accuracy of the models, but they fail by the strength of the results. This, in turn, underlines the necessity of applying advanced algorithms in those applications where a valid emotion recognition plays a major role [5],[18].

12.3 Discussion of the results and their implications, including an analysis of the challenges encountered and their impact on the overall research.

Results of the present study highlighted the strengths and limitations of the investigated machine learning algorithms for classifying affective states from EEG data. Each of these algorithms had its unique character, and their relative performances underlined some important observations concerning their practical applicability for performing emotion recognition tasks.

12.3.1 Advantages and Theoretical Perspectives

SVM and MLP Classifier revealed the best performance in both accuracy and reliability. The ability of SVM to define decision boundaries and the capability of MLP to model non-linear patterns in EEG data explain why these two models achieved relatively high values both for precision and recall in all investigated emotional states [13],[14]. This emphasizes their potential for applications requiring the highest classification accuracy in scenarios like monitoring mental health or adaptive human-computer interfaces.

That fact reflects their robustness against intrinsic noise and complexity associated with the nature of EEG data and, thus, is reliable in practical applications for real-world deployment [7],[18].

Logistic Regression indeed showed promising accuracy and proved especially good at classifying the NEUTRAL emotional state. However, its adherence to linear decision boundaries showed limitations in discriminating well between the NEGATIVE and POSITIVE classes [10],[18]. That implies that, although useful perhaps in pilot studies or when interpretability is a key issue, this method might not be satisfactory in tasks that require a finer discrimination of emotional states.

Its interpretable model structure makes the Decision Tree Classifier valuable for reasoning through decision-making processes. This classifier tends towards overfitting, however-as is evident in its misclassifications, particularly between the classes NEGATIVE and POSITIVE.

This underlines the trade-off between interpretability and performance, something quite common when working on high-dimensional EEG data [13],[18].

12.3.2 Model Overfitting and Generalization

The Decision Tree Classifier also appeared to overfit. While good performance was mostly seen on some splits, this resulted in high misclassification rates across cross-validation. This underlines the more general challenge of generalization in models, at least when subtle overlap exists between the classes of the dataset.

12.3.3 The Trade-off between Precision and Understandability

Although SVM and MLP Classifier had the highest accuracy, the non-interpretable nature of them raises several questions when it comes to such a critical domain as healthcare [7],[13]. At the same time, though more interpretable, Logistic Regression and Decision Tree lost in accuracy, therefore, a trade-off between these two factors needs to be considered according to application requirements [9],[10].

Optimal performance of SVM and MLP Classifier was obtained after an extensive and very computationally expensive hyper-parameter tuning [7],[14]. This points out a practical challenge of scalability of these methods at larger datasets or real-time applications [18],[21].

This comparison among these models further gives credence to the fact that model selection must be oriented toward the task at hand. For tasks requiring high accuracy and reliability, like emotion monitoring in clinical settings, SVM and MLP Classifier are obvious choices [16],[18]. In cases where model interpretability and explainability are important, though the performances of Logistic Regression and Decision Tree are relatively lower, those models might still be worth using.

13 Scalability of Machine Learning Models in EEG-Based Emotion Recognition

Electroencephalogram (EEG)-based emotion recognition is a rapidly evolving research area that utilizes machine learning (ML) approaches [13],[7]. Therefore, it is essential for these models to be scalable in more realistic situations, such as monitoring mental health, human-computer interaction, or adaptive learning systems [16],[18]. It covers what dictates scalability for ML models in the context of EEG-based emotion recognition, what the roadblocks are, and how they can be worked around.

13.1 What Affects Scalability

13.1.1 Data Volume and Quality

- **Volume:** The performance of ML models improves with more data [5],[21]. Collecting large, labelled datasets in EEG-based emotion recognition can be more challenging than in the classic domain, because not only a wide variety of participants but also a controlled environment is needed [9],[13].
- **Location of data:** Make sure to have clean data, less noise. Inter- and intra-subject variability in EEG signals from individual to individual or session to session can hinder model training and generalization [13],[18], [7],[14].

13.1.2 Model Complexity

- **Algorithm Selected**: The more algorithms selected to carry out a single task, the more scalable it becomes. They are trained on complex models such as deep neural networks (DNNs) that are good at capturing intricate patterns but tend to overfit without adequate computational power and training data [21],[22], [13],[23].
- **Feature Engineering:** Grab those effective techniques for feature extraction to improve performance on models. If these need a lot of preprocessing methods, that can scale [7],[14],[5],[18].

13.1.3 Computational Resources

- **Hardware Interdependence:** The training of larger models requires high-caliber hardware (GPU). The limited availability of such resources can limit the scalability of ML applications [23],[24].
- **Cloud Computing:** Cloud platforms can ease some resource limitations by offering on-demand scalable computing power [23], [24],[21].

13.1.4 Real-time Processing

- **Latency Requirements:** Numerous applications require real-time emotion detection, making low-latency processing capability an inherent need. This restriction can reduce the algorithm options and cause higher optimization needs [14], [18], [5], [21].
- **Streaming data handling:** The model must be able to effectively handle streaming data coming from the EEG devices, which adds to the complexity of model deployment and scalability [7],[14], [18],[23].

13.2 Challenges in Scalability

- **Inter-Subject Variability:** Variation in brain activity induced by different tasks for different groups can create an issue for transferability: a model trained on one subject group may not generalize to another. The wide variety means we need adaptive models or transfer-learning methods [7],[13].
- **Regulatory & Ethical Issues:** EEG data can bring about privacy challenges in addition to regulatory hurdles that can hinder the development of scalable solutions in enterprise systems [6],[17].
- **Multi-Modal Integration:** Integrating EEG data with other modalities, such as facial expressions or physiological signals, can improve emotion recognition accuracy while increasing model complexity and challenges in handling large datasets [23],[24].

13.3 Potential Solutions

- **Transfer Learning:** Leveraging pre-trained models from similar tasks can alleviate concerns over restricted data availability and inter-subject variability, allowing the method to be generalizable across populations [14],[18],[7].
- **Federated Learning:** One way to keep the data decentralized is by taking models trained over decentralized devices but keeping it local. It addresses privacy challenges and allows the use of open datasets without the risk of creating a centralized repository of sensitive information [6],[17].
- **Model Compression Techniques:** Pruning, quantization, and knowledge distillation are techniques used to reduce model size and complexity without sacrificing too much performance, making them more scalable for deployment on low-resource devices [5],[23].
- **Hybrid Models:** Hybridization of traditional machine learning methods and deep learning approaches could help reduce the amount of datasets and/or increase scalability across diverse problems [13],[18].
- **Incremental Learning Shapes:** It can be in incremental learning shapes, thus ensuring these models develop and consolidate variations/data shifts over several months or years [21],[14].

13.4 Conclusion

Various factors, like data volume, model complexity, computational resources, and real-time processing needs, affect the scalability of machine learning models in EEG-based emotion recognition. For real-world applicability, challenges like inter-subject variability and ethical implications have to be addressed. Researchers can expand the scalability of these systems for deploying them more widely in practical settings by using advanced techniques like transfer learning, federated learning, model compression, hybrid approaches, and continuous learning frameworks.

14 Ethical Considerations and Data Privacy

EEG-based emotion recognition does raise several ethical dilemmas and questions related to the privacy of individuals-all of which shall be considered with due care. By nature, EEG data is personal, and beyond capturing the emotional state of a person, it may even reveal information about one's cognitive state, his mental health, or even neural activities. This sensitive nature places greater responsibility on the researcher and developer to ensure stringent ethical standards while acquiring, storing, and using such data [6],[17]. Informed consent remains the cornerstone of any ethical data collection, and clear statements of what is collected and for what purposes should be given. This includes the clear explanation of data sharing, future research use, and the mechanisms put in place to protect participants from any potential misuse or exploitation [13].

The most important part of any ethical analysis of EEG is data security. Because of the risk of breaches, strong encryption and secure mechanisms of storage are required [17].

Complementing the technical measures should be the strategies of anonymization: this is, removal of identifiable features from the datasets. Such an approach will ensure that, even in cases of malicious access to the data, barring an association with individual contributors remains impossible [18]. What is more, guidelines of ethics should imply limited access to raw EEG data, allowing only authorized persons to perform analysis in certain controlled conditions. One of the serious ethical concerns is that emotion recognition technologies derived from EEG data would represent exploitation of people. Applications in ways such as employee monitoring, advertising, or criminal justice, without the subjects' effective consent, bring very serious ethical issues into play. Such applications risk inappropriate manipulations, violations of privacy, or even psychological traumas, since people might be secretly watched or classified according to their neurological pattern [17],[6].

It therefore needs comprehensive regulatory frameworks that can keep up with the acceptability of applications using EEG to recognize emotions in order to implement the deployment only in settings that contribute to the well-being and protection of basic rights [13],[18].

Other ethics concerns include reducing biases in the development of EEG-based systems. Due to demographics such as age, gender, and cultural background, neural signals might lead to prejudice and misclassification of emotions. Therefore, developers must create a balance in making datasets and pursue ways that reduce bias to ensure equity of performance across all populations. Open model limitation and bias reporting at every step of development create accountability and trust in the technology.

The potential for dual-use applications of EEG technologies further complicates the ethical landscape. While these systems can advance mental health care or enhance learning environments, on the other hand, they may be repurposed for coercive or manipulative practices. Developers and researchers carry a certain responsibility to predict and neutralize these risks by taking an ethical lead and making pro-active policy recommendations. In a nutshell, the ethical issues involved with EEG-based emotion recognition involve data privacy, consent, non-discrimination, and broader social ramifications of technologies [5],[13]. The effective response to these challenges should be developed within a multidisciplinary framework for the integration of robust technical protection in conjunction with open policy and commitment to ethical use of influential technology [6],[17]. Such measures will no doubt guarantee that EEG-based emotion recognition systems are bound to contribute toward the advance of human welfare by upholding respect for personal autonomy and privacy.

15 Conclusion

15.1 Summary of Findings and Contributions to the Field

This work systematically investigated the performance of the following four machine learning models for recognizing emotional states using EEG data: the Support Vector Machine, the Multi-Layer Perceptron Classifier, Logistic Regression, and the Decision Tree Classifier [7],[13],[18]. Such robust evaluations as K-Fold cross-validation and confusion matrix analysis were employed to find the best algorithms among the alternatives for this complex high-dimensional classification task [11]. The plots, therefore, very clearly reflected that the SVM and MLP Classifier were the top two performing models, achieving very excellent accuracy, precision, and recall on all three target classes: NEGATIVE, POSITIVE, and NEUTRAL. That means these models had captured the intricate nonlinear patterns in EEG data quite well and hence are more reliable and robust for real-world applications in emotion recognition [7],[16].

Logistic Regression, even with quite good accuracy, especially regarding the NEUTRAL class of emotional states' prediction, showed its limitation in distinguishing between NEGATIVE and POSITIVE classes [5],[9]. By its linear nature, it showed lack of capacity in grasping the fine relationship typically represented in EEG data, which can easily be accomplished by non-linear methods with huge success [13]. While similarly interpretable and easy to implement, the Decision Tree Classifier tended to overfit to the training data, as evidenced by higher cross-validation misclassification rates [10],[18]. This highlights the intrinsic tradeoffs between model interpretability and performance-a chief concern for both the researchers and practitioners in the field [5].

Another major result from this study was that visual methods of performance metrics, such as confusion matrices, provided complementary information about important numeric measures of performance [3],[8]. These visualizations provided a detailed view of how each model handled classification errors, which subtly showed performance differences not immediately apparent through accuracy alone. For example, the confusion matrices of the SVM and MLP Classifier showed how both models were steadfast in keeping false positives and false negatives as low as possible, especially on harder classes like NEUTRAL. These insights form the foundation of model selection that is both well-performing on average, as well as reliably so, for all emotional states [13],[18].

The contributions of this research go on to have many layers.

Firstly, the clear benchmark regarding the application of machine learning to EEG-based emotion recognition it provided was highly instructive guidance on the suitability of different algorithms for similar tasks [13],[18]. It points out that high-dimensional and noisy data, such as EEG signals, require strong preprocessing, like standardization of features. Thirdly, this study gives grounds to believe that the model complexity-interpretability trade-off is crucial in applications like mental health monitoring and adaptive human-computer interfaces, where transparency and reliability go side by side [6],[17].

It also touches on the problem at hand in some of the critical challenges in EEG-based emotion recognition, such as noisy data and generalization to small or various emotional states. This means that sophisticated models, such as the SVM and MLP Classifier, are effective in terms of accuracy but should be channeled toward the improvement of interpretability in further studies for wide applications in clinics and real-life scenarios [13],[16]. On the other hand, simpler models such as Logistic Regression and Decision Tree Classifier, while less accurate, might be more useful for ensemble methods or feature engineering to make strong improvements without paying a cost in terms of transparency [5],[10].

This work falls squarely within the emerging domain of affective computing, where neuroscience and AI integrate to bear enormous potential for innovative applications. This might provide an opening for further studies on personalized mental health care, emotion-driven interaction systems, and adaptive learning technologies using EEG-based emotion recognition through machine learning models. These results shall serve as a basis for more in-depth studies regarding algorithm optimization and the increase of datasets in order to further improve performance and applicability, narrowing the gap between theoretical research and practical deployment [13],[18],[16].

15.2 Future Research Directions and Model Enhancements

There are various avenues of further investigation and improvement of the model based on the result of this study. While the SVM and MLP Classifier achieved state-of-the-art performance in EEG data-based emotion classification, their reliance on a large amount of hyperparameter tuning and computation resource points toward scope for further optimization [7],[13],[18]. This could be referred to in future research using automated hyperparameter optimization techniques such as Bayesian Optimization or Grid Search, which make the process of model training smoother and more efficient with even better performance results [5],[23].

One promising direction in enhancement may be the application of ensemble methods, such as Random Forests, Gradient Boosted Trees, or Voting Classifiers [5],[18]. Ensemble methods somewhat compensate for the individual model weaknesses and work towards a more general classification. Moreover, combining with bagging or boosting will probably improve the generalization of even the most basic models, such as Logistic Regression and Decision Tree Classifier, and make at least them competitive with more advanced algorithms [10].

The second most important area for future work involves feature engineering and dimensionality reduction. EEG naturally carries high-dimensional and complicated data; therefore, further investigations with more sophisticated methods for feature selection, such as PCA or t-SNE, might reduce the noise and emphasize the most informative features. On top of all, domain-specific knowledge in neuroscience could be used, like feature extraction based on frequency or even connectivity analysis that better prepares the input features for emotion classification tasks [13],[15],[18].

Applications within the sphere of monitoring mental health to adaptive human-computer interactions have created an ever-increasing demand for more interpretable machine learning models [6],[16]. SVM and MLP Classifier had high accuracy, but interpretability remains an issue [7],[13]. Further research may involve developing more interpretable variants of these models or combining them with explanation techniques such as SHAP and LIME to better explain how the predictions are made [5].

Another important trend is the extension of the dataset and its diversification. So far, the results were obtained on one dataset, which is robust but cannot fully represent the variability of the EEG signal in different groups of subjects and under various conditions

[9]. Further studies could combine several datasets to develop a more generalized and representative model. Furthermore, the use of transfer learning techniques will most probably allow models, trained on one dataset, to be applied effectively to previously unseen data coming from other EEG datasets [14],[18].

The last point is that adding temporal dynamics might provide even better results. EEG signals represent time-series data, and the use of sequence-based models such as RNNs or Long Short-Term Memory networks could discover temporal patterns in data and provide further insight into how emotional states evolve over time [14],[18]. Hybrid models, which are combining spatial insights of traditional classifiers with the temporal strengths of sequential models, could represent the next step in this research.

In conclusion, automation in tuning and ensemble techniques must be incorporated; better feature engineering needs to come through in model performance. Also, interpretations and the ability for generalization need to be addressed. The expansion of datasets, exploration of temporal modelling, and integration of neuroscience-driven features will further bridge the gap between machine learning and practical, real-world applications, advancing the field of affective computing and EEG-based emotion recognition [6],[18].

16 Appendices

16.1 Appendix A: Dataset Details

Dataset Name: EEG Brainwave Dataset for Emotion Recognition

Source: Kaggle, EEG Brainwave Dataset: Feeling Emotions

The following dataset is a representation of EEG recordings that have been labeled based on their emotional states and is used to enable the process of research in emotion recognition. The details follow.

16.1.1 Overview of Dataset

Number of Rows (Samples): 2132

Number of Columns (Features): 2549

Label Distribution:

NEGATIVE: 932 samples

POSITIVE: 592 samples

NEUTRAL: 608 samples

```
Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2132 entries, 0 to 2131
Columns: 2549 entries, # mean_0_a to label
dtypes: float64(2548), object(1)
memory usage: 41.5+ MB
```

Figure 9 Dataset Information

16.1.2 Sample Data

Below is a snippet of the dataset for reference:

# mean_0_a	**mean_1_a**	**mean_2_a**	**mean_3_a**	**fft_747_b**	**fft_748_b**	**fft_749_b**	**label**
4.62	30.3	-356.0	15.6	-162.00	-162.00	280.00	NEGATIVE
28.80	33.1	32.0	25.8	-31.60	-31.60	2.57	NEUTRAL
8.90	29.4	-416.0	16.7	-148.00	-148.00	281.00	POSITIVE

```
Dataset Description:
        # mean_0_a     mean_1_a     mean_2_a     mean_3_a  ...     fft_746_b
count  2132.000000  2132.000000  2132.000000  2132.000000  ...  2132.000000
mean     15.256914    27.012462  -104.975629    13.605898  ...    95.104886
std      15.284621     9.265141   206.271960    16.874676  ...   203.194976
min     -61.300000  -114.000000  -970.000000  -137.000000  ... -1010.000000
25%       6.577500    26.075000  -195.000000     4.857500  ...    -8.837500
50%      14.100000    30.000000    14.950000    15.400000  ...    13.400000
75%      27.700000    31.400000    29.600000    26.500000  ...   149.250000
max     304.000000    42.300000   661.000000   206.000000  ...   888.000000

[8 rows x 2548 columns]
```

Figure 10 Dataset Description

The dataset is slightly imbalanced because the NEGATIVE class holds the majority of the samples. That is OK for supervised machine learning; however, one should pay extra attention to the balancing metrics when evaluating the performance.

16.1.3 Feature Description

The dataset contains a set of rich features that are extracted from EEG signals and can be grouped into:

Mean Values: Averaged values over EEG data for certain regions. For example, mean_0_a, mean_1_a.

Features Derived from Differences in Signal Values: mean_d_0_a, mean_d_1_a, etc.

Frequency Components: FFT values capturing frequency-domain information: fft_* columns.

Target Label : A categorical column, label, which is supposed to indicate the state of emotion: NEGATIVE, POSITIVE, or NEUTRAL.

16.1.4 Preprocessing, Standardization and Missing Values

Numerical data was pre-processed, standardizing features and handling any inconsistencies.

Standardize features by scaling all numeric features to have a mean of 0 and a standard deviation of 1 via Scikit-learn's StandardScaler.

There are no missing values in this dataset, hence, it is complete.

```
Missing Values in the Dataset:
# mean_0_a     0
mean_1_a       0
mean_2_a       0
mean_3_a       0
mean_4_a       0
              ..
fft_746_b      0
fft_747_b      0
fft_748_b      0
fft_749_b      0
label          0
Length: 2549, dtype: int64
```

Figure 11 Missing values in the Dataset

16.2 Appendix B: Code Implementation

This appendix provides a detailed exposition of the core computational components developed and utilized throughout the research. The modular scripts address key aspects of data preparation, model configuration, training, evaluation, and visualization, ensuring methodological transparency and reproducibility.

16.2.1 Data Preparation

The data_preparation.py script is designed to load and preprocess the dataset to ensure it is suitable for machine learning workflows. The following code illustrates the implementation.

```python
import pandas as pd
from sklearn.preprocessing import StandardScaler

def load_data(filepath): # Load the dataset
    data = pd.read_csv(filepath)
    print("First few rows of the dataset:")
    print(data.head())
    print("\nDataset Info:")
    print(data.info())
    print("\nDataset Description:")
    print(data.describe())
    print("\nMissing Values in the Dataset:")
    print(data.isnull().sum())
    return data

def preprocess_data(data): # Preprocess the dataset
    features = data.columns[:-1]
    scaler = StandardScaler()
    data[features] = scaler.fit_transform(data[features]) # Standardize the features
    return data

```

Figure 12 Data Preparation Script

16.2.2 Model Configuration

The model_config.py script defines the machine learning models deployed in this research. These models were selected based on their relevance to EEG-based emotion recognition. The script includes the following definitions.

```python
from sklearn.svm import SVC
from sklearn.neural_network import MLPClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier

def initialize_models(): # Initialize the models
    svm_model = SVC(kernel='linear', C=1.0) # Support Vector Machine
    ffnn_model = MLPClassifier(hidden_layer_sizes=(100,), activation='relu', solver='adam', max_iter=300) # Feedforward Neural Network
    logistic_model = LogisticRegression(solver='liblinear', penalty='l2', C=1.0) # Logistic Regression
    tree_model = DecisionTreeClassifier(max_depth=5, criterion='gini') # Decision Tree
    return svm_model, ffnn_model, logistic_model, tree_model
```

Figure 13 Model Configuration Script

16.2.3 Training and Evaluation

The train_evaluate.py script encapsulates the training and evaluation processes using K-Fold cross-validation. The modular design ensures robust performance evaluation for each model. Below are the key components.

```python
from sklearn.model_selection import KFold
from model_config import initialize_models
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

def train_test_models_k_fold(data, n_splits=5):
    """
    Trains and tests multiple models on the dataset using K-Fold cross-validation.

    Parameters:
    data (DataFrame): The input dataset with features and labels.
    n_splits (int): Number of folds for K-Fold cross-validation.

    Returns:
    results (dict): Dictionary containing model names and their mean accuracy scores.
    models (list): List of initialized models.
    """
    try:
        X = data.drop('label', axis=1)
        y = data['label']

        # Initialize models
        models = initialize_models()
        results = {}

        # Define the K-Fold cross-validator
        kf = KFold(n_splits=n_splits, shuffle=True, random_state=42)

        # Loop through each model for K-fold evaluation
        for model in models:
            model_name = type(model).__name__
            fold_accuracies = []

            for train_index, test_index in kf.split(X):
                X_train, X_test = X.iloc[train_index], X.iloc[test_index]
                y_train, y_test = y.iloc[train_index], y.iloc[test_index]

                # Train and evaluate the model on each fold
                model.fit(X_train, y_train)
                y_pred = model.predict(X_test)
                accuracy = accuracy_score(y_test, y_pred)
                fold_accuracies.append(accuracy)

            # Calculate and store the mean accuracy across folds
            results[model_name] = np.mean(fold_accuracies)
            print(f"{model_name} - Cross-validated accuracy: {np.mean(fold_accuracies):.2f}")

        return results, models
    except Exception as e:
        print(f"An error occurred: {e}")

def evaluate_model(X_train, X_test, y_train, y_test, model):
    """
    Evaluates a model using a confusion matrix and classification report on a single train-test split.

    Parameters:
    X_train, X_test, y_train, y_test: Training and testing data splits.
    model: The model to be evaluated.

    Returns:
    y_pred (array): The predicted labels.
    report (str): The classification report as a string.
    """
    try:
        # Train the model and generate predictions
        model.fit(X_train, y_train)
        y_pred = model.predict(X_test)

        # Generate and display the confusion matrix
        cm = confusion_matrix(y_test, y_pred, labels=["NEGATIVE", "POSITIVE", "NEUTRAL"])
        plt.figure(figsize=(8, 6))
        cm_df = pd.DataFrame(cm, index=["NEGATIVE", "POSITIVE", "NEUTRAL"], columns=["NEGATIVE", "POSITIVE", "NEUTRAL"])
        sns.heatmap(cm_df, annot=True, fmt='d', cmap='viridis', cbar=True)
        plt.xlabel('Predicted Label')
        plt.ylabel('True Label')
        plt.title('Confusion Matrix')
        plt.show()

        # Generate and print the classification report
        report = classification_report(y_test, y_pred, target_names=["NEGATIVE", "POSITIVE", "NEUTRAL"])


        return y_pred, report
    except Exception as e:
        print(f"An error occurred: {e}")
```

Figure 14 Training and Evaluation Script

16.2.4 Main Script

The main.py script integrates all components, orchestrating the workflow from data loading to model evaluation. The implementation is as follows.

```python
from data_preparation import load_data, preprocess_data
from train_evaluate import train_test_models_k_fold, evaluate_model
from sklearn.model_selection import train_test_split

def main():
    # Define the path to the dataset
    data_path = 'emotions.csv'

    # Load and preprocess the data
    print("Loading and preprocessing data...")
    data = load_data(data_path)
    data = preprocess_data(data)

    # Train and test models with K-Fold cross-validation for accuracy scores
    print("\nTraining and testing models with K-Fold cross-validation...")
    results, models = train_test_models_k_fold(data, n_splits=5)

    # Display cross-validated accuracy results for each model
    print("\nCross-validated model accuracy scores:")
    for model_name, score in results.items():
        print(f"{model_name}: {score:.2f}")

    # Perform a single train-test split for confusion matrix and classification report
    print("\nEvaluating models on a single train-test split...")
    X_train, X_test, y_train, y_test = train_test_split(data.drop('label', axis=1), data['label'], test_size=0.2, random_state=42)

    for model in models:
        print(f"\nResults for {type(model).__name__}:")
        y_pred, report = evaluate_model(X_train, X_test, y_train, y_test, model)

        # Print the classification report for the current model
        print("Classification Report:")
        print(report)

if __name__ == '__main__':
    main()
```

Figure 15 Main Script

16.2.5 Visualization

The visualizations.py script supports the graphical representation of model evaluation results, focusing on confusion matrix visualization. The implementation is as follows.

```python
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.metrics import confusion_matrix

def plot_confusion_matrix(y_true, y_pred, title='Confusion Matrix'):
    cm = confusion_matrix(y_true, y_pred)
    sns.heatmap(cm, annot=True, fmt='d')
    plt.xlabel('Predicted')
    plt.ylabel('True')
    plt.title(title)
    plt.show()
```

Figure 16 Visualization Script

17 References

[1] Birdy654, "EEG Brainwave Dataset: Feeling Emotions," Kaggle, 4 December 2024. [Online]. Available: https://www.kaggle.com/datasets/birdy654/eeg-brainwave-dataset-feeling-emotions. [Accessed 4 December 2024].

[2] Y. B. a. A. C. Ian Goodfellow, Deep Learning, MIT Press, 2016.

[3] F. P. e. al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research,* 2011.

[4] S. R. a. P. Norvig, Artificial Intelligence: A Modern Approach, Pearson, 2020.

[5] R. T. a. J. F. Trevor Hastie, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, 2009.

[6] J. M. e. al., "The Ethics of Algorithms: Mapping the Debate," *Big Data & Society,* 2016.

[7] R. H. Y. W. a. X. Z. D. Liu, "EEG-Based Emotion Classification Using Deep Learning Networks," *Frontiers in Neuroscience,* 2020.

[8] s.-l. contributors, "Scikit-learn: Machine Learning in Python," Scikit-learn, 2011. [Online]. Available: https://scikit-learn.org. [Accessed December 2024].

[9] S. K. e. al., "DEAP: A Database for Emotion Analysis Using Physiological Signals," *IEEE Transactions on Affective Computing,* no. IEEE, 2012.

[10] J. R. Quinlan, "Induction of Decision Trees," *Machine Learning,* 1986.

[11] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," in *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*.

[12] M. G. T. a. D. I. F. A. T. Tzallas, "Epileptic Seizure Detection in EEGs Using Time–Frequency Analysis," *IEEE Transactions on Information Technology in Biomedicine,* 2009.

[13] S. M. A. a. M. J. Fonseca, "Emotions Recognition Using EEG Signals: A Survey," *IEEE Transactions on Affective Computing,* 2017.

[14] I. R. M. Y. a. N. C. P. Bashivan, "Learning Representations from EEG with Deep Recurrent-Convolutional Neural Networks," in *Proceedings of the 2015 IEEE International Conference on Pattern Recognition (ICPR)*, 2016.

[15] R. J. C. a. R. J. Barry, "Time-Frequency Analysis of EEG in Cognitive Neuroscience," *International Journal of Psychophysiology,* 2000.

[16] J. A. a. D. Campos, "Improving BCI-Based Emotion Recognition by Combining EEG Feature Selection and Kernel Classifiers," *Expert Systems with Applications,* 2016.

[17] M. D. R. a. C. S. Economou, "Privacy Concerns in the Collection and Use of Brain Data," *Ethics in AI Research Journal,* 2020.

[18] S. K. a. M. G. T. Ince, "EEG Feature Extraction and Classification Using Wavelet Transform and Machine Learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* 2016.

[19] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," in *Computing in Science & Engineering*, 2007.

[20] OpenAI, "ChatGPT: An AI Language Model," OpenAI, [Online]. Available: https://chat.openai.com/.

[21] T. M. Mitchell, Machine Learning, McGraw-Hill.

[22] C. M. Bishop, Pattern Recognition and Machine Learning, New York, NY, USA: Springer.

[23] J. Schmidhuber, "Deep Learning in Neural Networks: An Overview," Neural Networks Journal, [Online]. Available: https://doi.org/10.1016/j.neunet.2014.09.003.

[24] Y. B. a. G. H. Y. LeCun, "Deep Learning," Nature Publishing Group, [Online]. Available: https://doi.org/10.1038/nature14539.

[25] A. C. Management, "Pandas," [Online]. Available: https://pandas.pydata.org/.

[26] M. L. Waskom, "seaborn: statistical data visualization," Journal of Open Source Software, [Online]. Available: https://doi.org/10.21105/joss.03021..

[27] C. M. K. v. d. W. S. Harris, "Array programming with NumPy," [Online]. Available: https://www.nature.com/articles/s41586-020-2649-2.

[28] J. D. Hunter, "NumPy," [Online]. Available: https://numpy.org/.

[29] J. D. Hunter, "Matplotlib: A 2D graphics environment," Computing in Science & Engineering, [Online]. Available: https://matplotlib.org/.