A NOVEL HYBRID R-CNN-TPOT MODEL FOR PREDICTING BRAIN AGE FROM EEG DATA

Anupallavi S, Rutwik, Shreya Kamalapurkar, Sanjay P

Department of Artificial Intelligence and Machine learning, Acharya Institute of Technology, Bangalore, India

ABSTRACT

With advancing age, significant transformations in electroencephalogram (EEG) signals are evident, which can be defined as brain age. One can identify discrepancies from conventional aging patterns by juxtaposing chronological age with brain age. Although Machine Learning (ML) techniques have been broadly applied in predicting brain age through brain imaging, EEG-based strategies have not been sufficiently explored, particularly about the Tree-based Pipeline Optimization Tool (TPOT). This study proposes a new hybrid ML technique for estimating brain age utilizing EEG data to fill this void. The approach integrates various features, including spectral, statistical, frequency-domain, and decomposition-domain attributes. Moreover, a novel model called Regression-based Convolutional Neural Network-TPOT (R-CNN-TPOT) has been developed for this analysis. This model merges the mathematical framework of CNN with the TPOT classifier via regression modelling. The performance of R-CNN-TPOT was measured using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), R-squared, and Root MSE (RMSE), achieving values of 0.033, 0.063, 0.095, and 0.251, respectively.

KEYWORDS

Brain Age Estimation, EEG, TPOT, Convolutional Neural Networks, Regression Modelling

1. INTRODUCTION

Across the globe, millions of individuals encounter challenges associated with the Central Nervous System (CNS) as they grow older, which imposes a considerable strain on national healthcare systems and caregivers[1]. The transition from childhood to adolescence is marked by significant transformations in neural brain tissues, including processes such as synaptic pruning, myelination, synaptogenesis, and dendritic arborization. These changes lead to modifications in Gray matter and an overall increase in brain volume[2]. As individuals age, both functional and structural connectivity within the brain become more pronounced, especially among various regions and across hemispheres [3]. Functional connectivity (FC) describes the spatial and temporal correlations within brain networks during periods of rest and activity, typically experiencing a decline after approximately 65 years of age as part of the normal aging process [4]. In recent decades, extensive research has been dedicated to understanding the brain's aging process, with a growing interest in the role of genetics [5].

In recent years, there has been a growing interest in estimating biological aging. Biological age is influenced by a variety of factors, including epigenetic markers, mitochondrial performance, telomere length, and cellular properties. Machine learning (ML) techniques can correlate brain

DOI:10.5121/cseij.2025.15121

data with chronological age, serving as a surrogate for biological age [8] Among these, EEG signals are particularly noteworthy due to several benefits: they are affordable, non-invasive, and can be recorded in a home setting; they capture functional changes rather than structural ones; and they facilitate repeated assessments within the same individual to measure the effectiveness of interventions such as medication or brain stimulation, which aim to maintain or improve cognitive function and longevity [7]. Additionally, EEG biomarkers are effective in distinguishing between typically developing children and those with developmental disorders, including autism and ADHD [9].

Numerous studies have investigated age-related changes in the brain through imaging modalities such as Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), and EEG, with the latter being one of the most cost-effective and accessible techniques available [10]. In contrast to conventional statistical methods, ML models excel at making inferences on an individual basis rather than at the group level, thereby enhancing their utility in clinical settings [11]. Furthermore, Automated Machine Learning approaches have gained prominence due to their proficiency in managing complex large datasets, automating the selection of the most suitable techniques, and optimizing hyperparameters, which ultimately leads to enhanced predictive accuracy and dependability. TPOT, in particular, stands out for its ability to automate the creation and optimization of ML workflows through Genetic Programming (GP), evolving computer programs tailored to specific problem domains without necessitating human involvement [12].

This research presents an innovative hybrid machine learning approach for estimating brain age using EEG signals. The study employs the HarMNqEEG dataset for this purpose. Initially, the EEG signals are pre-processed to eliminate artifacts through the application of a Gaussian filter. A range of features, including spectral, statistical, frequency-domain, and decomposition-based characteristics, are extracted from the cleaned EEG signals. These features are then integrated and analysed through a feature selection process utilizing the Canberra distance metric to pinpoint the most significant features. The chosen features are then fed into the R-CNN-TPOT model for predicting brain age.

- This study makes several key contributions, including the following:
- The development of the R-CNN-TPOT model for brain age prediction represents a novel hybrid machine learning approach that utilizes EEG signals. This model combines the convolutional neural network (CNN) framework with the TPOT classifier, employing regression techniques grounded in the Fractional Concept (FC).

The structure of the paper is organized as follows: Section 2 reviews current methods for predicting brain age, Section 3 details the R-CNN-TPOT model introduced in this research, Section 4 analyses the experimental results, and Section 5 offers concluding remarks.

2. MOTIVATION

Individuals with neuroanatomical disorders often exhibit a mismatch between their chronological age and the development of their brain, where brain age acts as an essential biomarker for evaluating maturity levels linked to behavioural changes. Given that EEG can detect structural changes in the brain, it offers a promising approach for estimating brain age. This section explores the current methodologies for predicting brain age using EEG signals, which led to the creation of the R-CNN-TPOT model.

2.1. Literature Review

Beheshti, Iman and colleagues [2] developed a method for predicting brain age using machine learning (ML) techniques that analyse EEG signals. They utilized three classifiers: Random Forest (RF), Relevance Vector Machine (RVM), and Support Vector Machine (SVM) to estimate brain age from pre-processed EEG data. Among these classifiers, RVM demonstrated the best performance. The method produced consistent results and was effective even with limited input data; however, issues such as significant collinearity and inadequate feature randomization in RVM resulted in increased error variance due to obscured features. Cook., et al. [13] proposed a brain age prediction method that utilized raw EEG signals, employing machine learning models including Bidirectional Gated Recurrent Unit (BGRU), Bidirectional Long Short-Term Memory (BLSTM), GRU, and LSTM. BGRU achieved the most favourable results by directly learning from raw data without preprocessing, although its learning rate significantly decreased as the number of epochs increased. Jeong, Hyerin., et al. and colleagues [9] developed a region-specific brain age prediction method using machine learning, which incorporated a tree-based feature selection process to identify a minimal set of features. Machine learning attained high predictive accuracy even with a limited number of features, although the study did not explore the potential advantages of including eight brain lobes, which could have enhanced the model's performance Al Zoubi., et al. [6] implemented a Machine learning and stack-ensemble learning approach for brain age prediction based on EEG signals. After preprocessing the EEG data and extracting various features, they eliminated highly correlated and low-variance features using NCV, and then applied stack-ensemble learning for prediction. While this method effectively captured informative features and the spatial distributions of key predictors, it was marked by high computational complexity.

Tanveer, Muhammad., et al. [10] introduced a method for predicting brain age through the use of Deep Convolutional Neural Networks (DCNN). Their methodology involved an initial denoising of EEG signals through automatic preprocessing, followed by data augmentation to increase the sample size. The DCNN was then employed to predict brain age by analysing segments of the EEG signal. Although this method was effective in extracting pertinent information for predictions, it encountered difficulties in ensuring the generalizability of its findings. Baecker., et al. [4] created a brain age prediction technique that utilized ML models TPOT, relying on Magnetic Resonance Imaging (MRI) data. The MRI data underwent preprocessing, after which TPOT was used to identify the optimal machine learning pipeline. The final predictions were made using RVR, with both TPOT and RVR being fine-tuned on the training dataset. This approach yielded reliable results without the need for extensive system adjustments, although it faced challenges regarding feature interpretability. Sun, H., et al. [7] presented an interpretable machine learning method for estimating brain age from sleep EEG signals. Engemann, D.A., et al. [1] predicted brain age by integrating deep learning and traditional machine learning techniques applied to Magnetoencephalography (MEG) and EEG (M/EEG) signals. They established a benchmark dataset for brain age prediction, preprocessing the M-EEG data using the MNE-BIDS pipeline and employing various machine learning and deep learning models. This approach exhibited strong generalizability and produced results with low computational costs, but it struggled to effectively utilize supervised learning to combine multiple averaging methods for improved prediction outcomes.

3. CHALLENGES

The issues faced by the available approaches for brain age prediction utilizing EEG signals are presented below.

- The machine learning approach employing RVM, as presented in [2], proved ineffective in leveraging EEG signals to establish the connection between behavioural phenotypes and developmental stages for forecasting attention issues and academic performance.
- The study in [13] employed the BLSTM model for estimating brain age from EEG signals, but it fell short in utilizing signal preprocessing and various extraction methods to boost the precision of the brain age predictions.
- The RF-based method outlined in reference [9] exhibited strong predictive capabilities; however, it failed to account for establishing an appropriate range for the predicted brain age during the prediction process, as variations in the distribution of predicted brain age are noted with aging.
- A significant concern identified in the NCV and the stacked ensemble learning method outlined in [6] was that it had not been utilized to forecast other pertinent physiological metrics derived from EEG signals of the brain.

Although several machine learning methods have been established for predicting brain age based on EEG signals, the restricted number of EEG samples limits the models' effectiveness in generalizing patterns, which poses significant challenges for accurate brain age prediction.

The prediction of brain age is crucial for assessing the neural activities within the brain, particularly in individuals suffering from neurodegenerative disorders. This section outlines the various methodologies implemented in the development of the R-CNN-TPOT framework proposed in this study for brain age estimation. The prediction process utilizes EEG signals sourced from a comprehensive database. Initially, the input EEG signals undergo pre-processing, where a Gaussian filter [14] is employed to eliminate noise. Following this, a range of features is extracted from the signals, including spectral-based features, statistical features, frequency domain features, and decomposition domain features. The spectral features analysed encompass the Amplitude Modulation Spectrogram (AMS) [15], spectral flux, tonal power ratio, Spectral Centroid (SC) [16], spectral spread, Power Spectrum Density (PSD) [17], and Logarithmic Band Power (LBP) [17]. The statistical features considered include mean, median, standard deviation, kurtosis, skewness, and entropy [18]. Additionally, frequency domain features such as Fourier transform and band power [19], along with decomposition domain features like Empirical Mode Decomposition (EMD) and Discrete Wavelet Transform (DWT), are derived from the cleaned EEG signals. Subsequently, all extracted features are integrated, and the most relevant features are selected using the Canberra distance measure [20]. This process involves calculating feature correlations based on the Canberra distance, allowing for the removal of redundant features to enhance the quality of the feature subset. The refined features are then utilized in the brain age prediction module, where the R-CNN-TPOT, a novel integration of the CNN model [21] and the TPOT classifier through regression modelling, is employed for the prediction task. A block diagram illustrating the proposed R-CNN-TPOT for brain age prediction using EEG signals is presented in Figure 1.

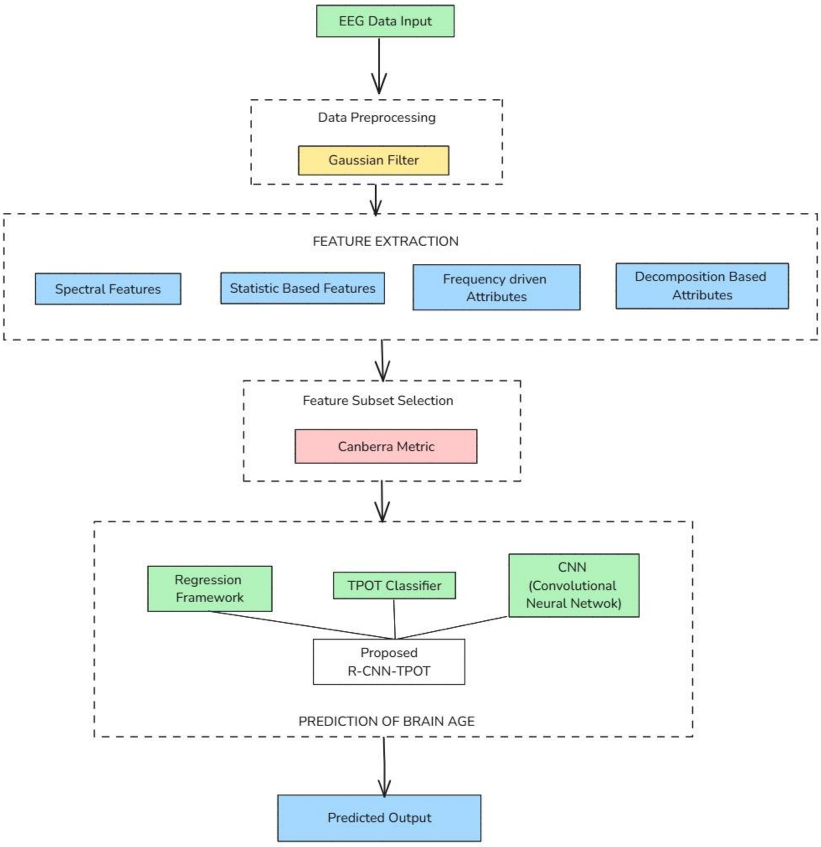

Figure 1. Block diagram of the proposed R-CNN-TPOT for brain age prediction using EEG signal

3.1. Data Acquisition

The EEG signals utilized for estimating brain age are derived from the dataset, which is referenced as, .

$$A = \{A_1, A_2, .., A_k, .., A_a\} \tag{1}$$

where, A_k characterizes the k^{th} EEG record taken from the dataset A which contains a total of a records, and the EEG signal A_k is employed in brain age prediction.

3.2. Signal Processing

EEG signals are captured from the brain's electrical activity, but the recording process can be affected by various artifacts, including both physiological and non-physiological types, which may obscure the original brain signal. Since the brain signal is relatively weak, it is important to implement denoising methods on the EEG signals. In this case, a Gaussian filter [14] is utilized for this purpose, as its Fourier transform is a Gaussian function that is cantered at zero frequency. The EEG signal is modified through convolution with a Gaussian function, leading to a refined output. The Gaussian filter is mathematically represented by the following expression.

$$\vartheta(A_k) = \frac{1}{\sqrt{2\pi} \cdot sd} e^{\frac{A_k^2}{2sd^2}} \tag{2}$$

where, sd refers to the Standard Deviation (SD), and A_k is the input EEG signal. Consider the pre-processed EEG signal obtained to be denoted as b and this is applied to the next module.

3.3. Feature Extraction

The process of feature extraction is centered on reducing redundancy in information while ensuring that the most relevant features of the EEG signal are retained. This involves utilizing the pre-processed EEG signal to derive a variety of features, which include spectral-based, statistical, frequency domain, and decomposition domain features. The statistical features considered consist of mean, median, standard deviation, kurtosis, skewness, and entropy [18]. Furthermore, frequency domain features such as Fourier transform and band power [19], as well as decomposition domain features like empirical mode decomposition (EMD) and discrete wavelet transform (DWT), are extracted from the denoised EEG signal. A more detailed discussion of these features will follow in subsequent sections.

3.3.1. Spectral-Based Features

The analysis of the EEG signal's spectral characteristics relies on spectral-based features. In this process, several features, such as AMS, spectral flux, tonal power ratio, spectral centroid, spectral spread, are extracted from the pre-processed EEG data.

i) AMS

AMS is utilized for the analysis of EEG signals by decomposing them into a three-dimensional representation that aligns with modulation across the modulation-frequency and time-frequency axes.

ii) Spectral flux

The extraction of spectral components from the EEG relies on spectral flux [22], which is vital because changes in these components can impair the system's efficiency. Spectral flux is defined as the squared difference of the normalized magnitudes of the spectral distributions associated with successive EEG frames.

iii) Spectral centroid (SC)

SC [16] is assessed by evaluating the "centre of gravity" based on the frequency and magnitude data from the Fourier transform. This can be articulated as the proportion of the frequency-weighted amplitude value relative to the total cumulative amplitude value.

iv) Tonal power ratio

This feature [22] is utilized to assess the tonal characteristics of the EEG signal, calculated by determining the ratio of the tonal power of the spectral components to the overall power

v) Spectral spread

This feature [16] quantifies the extent to which the spectrum diverges from SC (Standard Curve), and is also known as the second central moment.

3.3.2. Statistical Features

In addition, different statistical properties, such as mean, median, standard deviation, kurtosis, skewness, and entropy, are derived from the pre-processed EEG signal.

i) Mean

The mean represents the average amplitude value derived from various samples within the EEG signal, and it is expressed as follows:

ii) Median

The median is established by finding the central value of the EEG sample that has been arranged in order, determined as follows.

iii) SD

This parameter serves to measure how much the sample value varies from the mean.

iv) Kurtosis

This feature serves to measure the extent of deviation in the tail of the pre-processed EEG signal distribution compared to the normal distribution.

v) Skewness

The term skewness describes the level of asymmetry of the EEG signal concerning its average value.

vi) Entropy

An essential attribute of EEG signals is entropy, which reflects the non-linear aspects embedded in the signals.

3.4. Feature Selection

Once the prominent features have been derived from the pre-processed EEG signal, the feature vector is processed in the feature selection phase. Here, the most suitable features are selected based on Canberra distance [20].

$$CD = \sum_{l=1}^{f} \frac{|x_l - y_l|}{|x_l| + |y_l|} \tag{3}$$

where, CD refers to the Canberra distance x refers to the candidate feature, and y characterizes the target feature. The prominent features are thus selected and the reduce feature set obtained is symbolized as Z_{fs}, which is applied to the R-CNN-TPOT for brain age prediction.

3.5. Brain Age Prediction

This research employs a hybrid network known as R-CNN-TPOT to predict brain age, leveraging EEG signal inputs and a reduced set of features. The R-CNN-TPOT integrates the mathematical principles of CNNs [21] with the TPOT classifier, utilizing regression modeling techniques.

3.5.1. Proposed R-CNN-TPOT

The R-CNN-TPOT framework consists of three primary modules: the TPOT classifier, the R-CNN-TPOT layer, and the CNN model. The TPOT classifier utilizes a decision tree [25] based on a tree structure. The process begins with the input EEG signal being analyzed by the TPOT classifier, which predicts the brain age using the decision tree. A schematic diagram of the R-CNN-TPOT is shown in Figure 2, with detailed explanations of the three modules and their functions provided in the following sections.

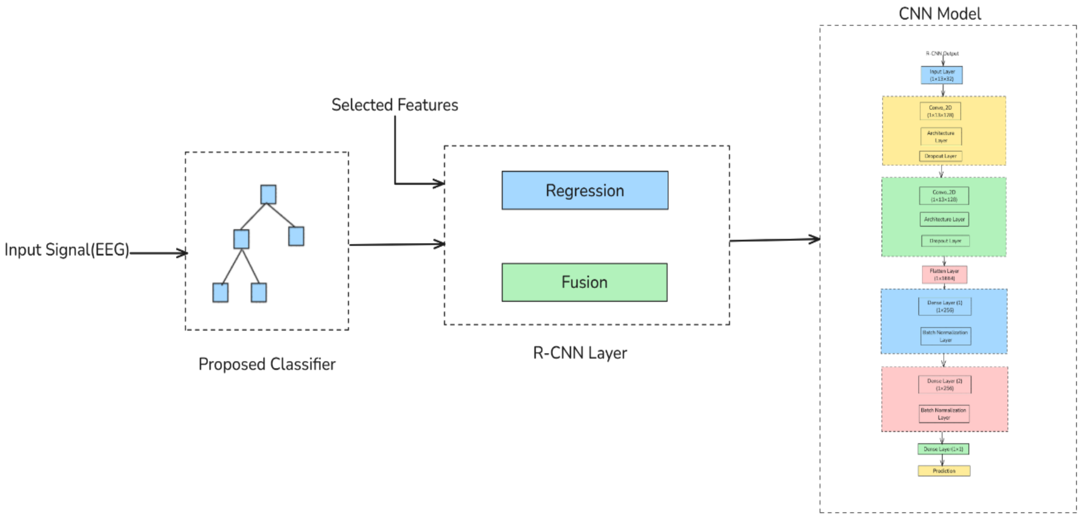

Figure 2. Architectural view of the R-CNN-TPOT

a) TPOT

The TPOT [12] framework utilizes Genetic Programming (GP) as its foundation for an AutoML system aimed at improving the accuracy of supervised classification processes. TPOT is particularly valued for its scalability, ease of use, and the generation of high-quality pipelines.

b) TPOT classifier using regression modelling

This analysis considers a tree structure inspired by decision trees, with the input EEG signal being processed through the decision tree for predictive analysis. Decision trees are a widely recognized machine learning approach, classified as a supervised learning technique, and are often used to address regression and classification tasks.

a) Construction level

A top down approach is utilized for constructing the decision trees by taking into account the training set $U = \{P_1, P_2, .., P_\alpha\}$, where P represents the objects contained in the training set and the training set has a total of α objects. Let T refers to the subset of P , and $S_i, i:1:u$ denotes the set of u attributes. Every attribute S_i can have a value from the set ψ^{S_i}, which contain all κ potential that S_i can have. Here. $\psi = \{\psi_1,..,\psi_H\}$ indicates the potential H output classes. Further, $Y = \{Y_1,...,Y_v\}$ indicates the v leaf nodes produced while constructing the decision tree.

b) Classification level

The objects are classified based on the attributes created with Basic Belief Assignments (BBAs). Let $S = \{S_1,...,S_u\}$ designates the u attributes representing the testing cases, and the cumulative count of testing cases $P_k; (k = \{1,...,\alpha'\})$ be represented as α'. Let ψ^S denote the global discernment frame associated with every attribute and is obtained as follows,

$$\psi^S = \underset{i=1,..,u}{\times} \psi^{S_i} \tag{4}$$

The value of ψ^S is obtained by taking the cross product of all other ψ^{S_i}. Thereafter, the combined BBAs of all objects showcasing its belief on the attribute value is computed. This is carried out by performing the following.

c) R-CNN-TPOT

The output C_1 generated by the TPOT is subsequently applied to the R-CNN-TPOT layer, where it is combined with the reduced feature set Z_{fs} based on regression modelling. The fused output produced by this layer is then subjected to the CNN module for making the final prediction. FC [26], a branch of applied mathematics that utilizes Laplace transform for finding solution to any complex integral and derivative function. At time t, the output of the R-CNN-TPOT is regarded to be obtained based on the reduced feature set Z_{fs}, and this is modelled as,

$$D = \sum_{\delta=1}^{n} Z_{fs}^{\delta} * \omega_{\delta} \tag{5}$$

where, ω stipulates the weight, Z_{fs} is the selected feature set, and D is the output at the time interval t, and at $t-1$, the output is based on the output of the TPOT C_1. These can be combined using FC, and applying FC [26],

$$C_2 = \beta D + \frac{1}{2}\beta * C_1 \tag{6}$$

Substituting the value of D from equation (24),

$$C_2 = \beta \sum_{\delta=1}^{n} Z_{fs}^{\delta} * \omega_{\delta} + \frac{1}{2}\beta * C_1 \tag{7}$$

Here, β is a constant, and the fused output C_2 is applied to the CNN module for getting the final predicted value of brain age.

4. RESULTS AND DISCUSSION

In this segment, we review the experimental findings from the application of the R-CNN-TPOT for brain age prediction, showcasing the effectiveness of the proposed approach. Furthermore, a detailed account of the experimental configuration and the assessment methodology is presented.

4.1. Experimental set-up

The R-CNN-TPOT developed for estimating brain age from EEG signals in this study is executed on a system utilizing the Python programming language.

4.2. Dataset Description

The EEG data for this analysis is derived from the Harmonized-Multinational qEEG norms (HarMNqEEG) dataset [27] [28]. The dataset initially included 1792 samples, which were subsequently narrowed down to 1564 samples after conducting a quality review, with 781 samples from males and 783 from females. EEG data was collected from nine countries, namely the USA, Switzerland, Russia, Malaysia, Germany, Cuba, Colombia, China, and Barbados, encompassing individuals aged between 6 and 95 years.

4.3. Evaluation measures

To measure the effectiveness of the R-CNN-TPOT, metrics like MSE, R-squared, and RMSE are taken into account.

MAE: This parameter calculates the mean value of the absolute differences between the predicted brain age and the actual brain age.

ii) MSE: The MSE serves as a metric to assess the average squared deviation between the predicted brain age and the actual brain age.
iii) R-squared error:This parameter represents the degree of correspondence between the predicted brain age and the actual brain age, also known as the Coefficient of Determination.
iv) RMSE: The RMSE parameter is derived by taking the square root of the MSE value.

4.4. Performance Assessment

The performance evaluation of the R-CNN-TPOT is conducted using various parameters in relation to different training sets and epochs, as illustrated in figure 3. In figure 3a), the performance of the R-CNN-TPOT is assessed through Mean Absolute Error (MAE). For an 80% training set, the MAE values recorded by the R-CNN-TPOT are 0.760, 0.560, 0.356, and 0.092 for 10, 20, 30, and 40 epochs, respectively. Figure 3b) further explores the R-CNN-TPOT's performance in terms of Mean Squared Error (MSE), showing MSE values of 0.687, 0.556, 0.383, and 0.080 for the same epochs with the 80% training set. Additionally, figure 3c) presents an analysis based on R-squared values, where the R-CNN-TPOT achieved scores of 0.656 for 10 epochs, 0.445 for 20 epochs, 0.355 for 30 epochs, and 0.085 for 40 epochs, all with an 80% training set. Lastly, figure 3d) illustrates the evaluation of the R-CNN-TPOT using Root Mean Squared Error (RMSE), with values of 0.829, 0.746, 0.619, and 0.283 for 10, 20, 30, and 40 epochs, respectively, again utilizing the 80% training set.

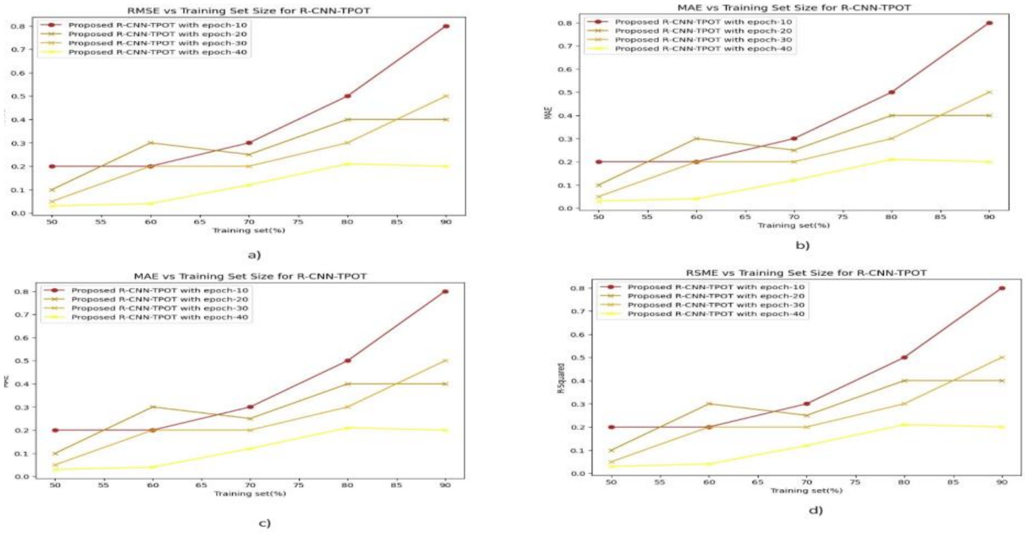

Figure 3. Performance evaluation of the R-CNN-TPOT based on a) MAE, b) MSE, c) R-squared, and d) RMSE

4.5. Comparative Assessment

An assessment of the R-CNN-TPOT is conducted to determine its dominance over other techniques by modifying the training set parameters and the k-value.

4.5.1. With Varying Training Set

he assessment of the R-CNN-TPOT was conducted using different training sets, as illustrated in Figure 4. In Figure 4a), the performance of the R-CNN-TPOT is analyzed in terms of Mean Absolute Error (MAE). The R-CNN-TPOT achieved an MAE of 0.365 with an 80% training set, whereas other brain age prediction methods, including RVM, BGRU, RF, and NCV, along with stacked ensemble learning, recorded higher MAE values of 13.279, 10.651, 6.276, and 6.170, respectively. Figure 4b) presents the evaluation of the R-CNN-TPOT based on Mean Squared Error (MSE). For the 80% training set, the MSE values for RVM, BGRU, RF, NCV, and stacked ensemble learning were 12.377, 9.474, 5.206, 3.178, while the R-CNN-TPOT achieved a significantly lower MSE of 0.320. Additionally, Figure 4c) depicts the analysis of the R-CNN-TPOT in terms of R-squared values. The R-CNN-TPOT recorded an R-squared value of 0.084, outperforming RVM, BGRU, RF, and NCV, which had values of 14.847, 13.981, 12.714, and 7.402, respectively, for the 80% training set. Finally, Figure 4d) illustrates the investigation of the R-CNN-TPOT concerning Root Mean Squared Error (RMSE). With an 80% training set, the RMSE values were 3.518 for RVM, 3.078 for BGRU, 2.282 for RF, 1.783 for NCV and stacked ensemble learning, while the R-CNN-TPOT achieved an RMSE of 0.566.

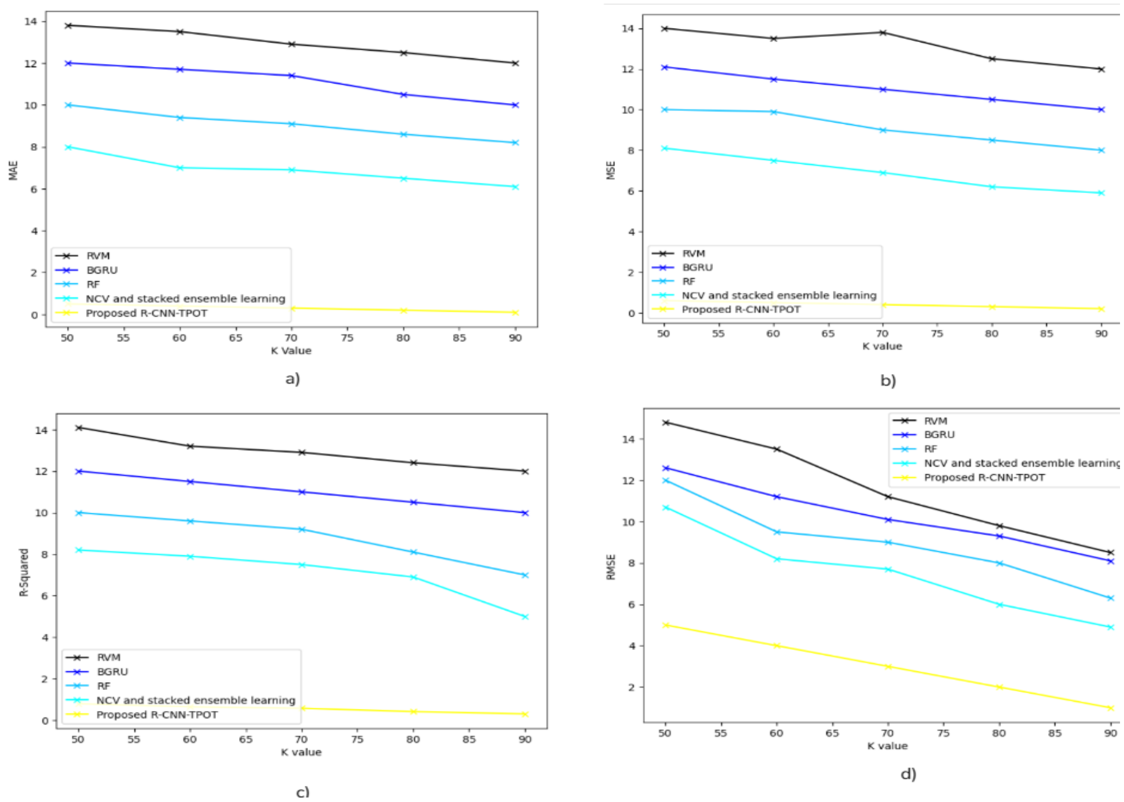

Figure 4. Comparative evaluation of the R-CNN-TPOT based on a) MAE, b) MSE, c) R-squared, and d) RMSE with varying training set

4.5.2. Based on K-Value

Figure 5 illustrates the assessment of the R-CNN-TPOT across various performance metrics with different K-values. In panel 5a), the analysis of the R-CNN-TPOT based on Mean Absolute Error (MAE) is presented. At a k-value of 8, the MAE values for RVM, BGRU, RF, NCV, stacked ensemble learning, and R-CNN-TPOT are recorded as 12.268, 10.524, 8.386, 5.089, and 0.379, respectively. Additionally, panel 5b) showcases the evaluation of the R-CNN-TPOT in terms of Mean Squared Error (MSE). The R-CNN-TPOT achieved an MSE of 0.270 at a k-value of 8, while RVM, BGRU, RF, NCV, and stacked ensemble learning reported higher MSE values of 12.721, 9.375, 7.353, and 2.045, respectively. Panel 5c) depicts the R-squared assessment of the R-CNN-TPOT, where the existing models recorded R-squared values of 14.266 for RVM, 13.024 for BGRU, 10.751 for RF, and 6.466 for NCV and stacked ensemble learning, while the R-CNN-TPOT achieved a significantly lower R-squared error of 0.081 at a k-value of 8. Finally, panel 5d) presents the evaluation of the R-CNN-TPOT in terms of Root Mean Squared Error (RMSE). The RMSE for the R-CNN-TPOT is measured at 0.520, whereas the other models, including RVM, BGRU, RF, and NCV, along with stacked ensemble learning, reported higher RMSE values of 3.567, 3.062, 2.712, and 1.430, respectively, at a k-value of 8.

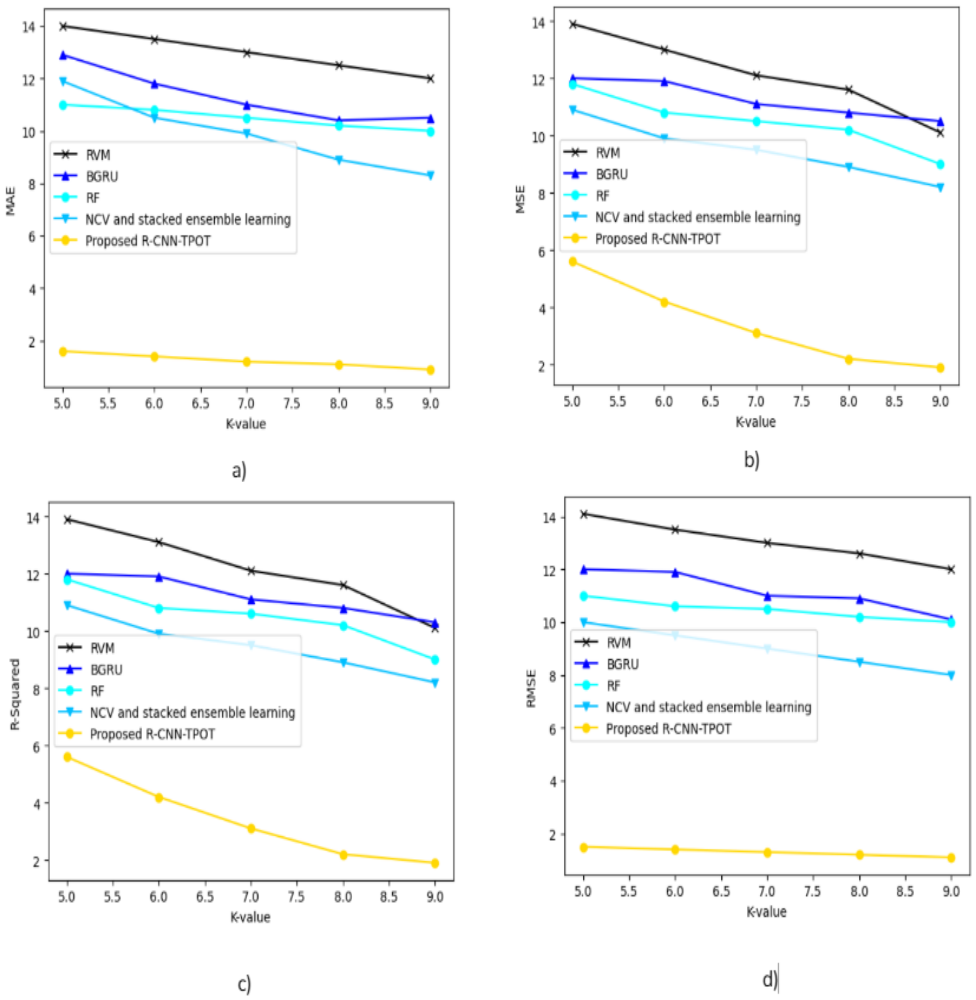

Figure 5. Comparative evaluation of the R-CNN-TPOT based on a) MAE, b) MSE, c) R-squared, and d) RMSE with respect to K-value

4.6. Comparative Discussion

This research introduces a novel hybrid machine learning approach aimed at predicting brain age using EEG signals, with a detailed comparison of the R-CNN-TPOT method presented in this section. Table 1 illustrates the performance metrics obtained during the experiments, which were conducted with a training set size of 90% and a k-value of 9. Notably, the R-CNN-TPOT achieved a minimal mean absolute error (MAE) of 0.033, significantly lower than the MAE values of 11.496, 10.005, 7.141, and 4.920 recorded by existing brain age prediction methods such as RVM, BGRU, RF, and NCV, as well as stacked ensemble learning. Additionally, the R-CNN-TPOT demonstrated a low mean squared error (MSE) of 0.063, in contrast to the MSE values of 11.602 for RVM, 7.566 for BGRU, 4.674 for RF, and 1.485 for NCV and stacked ensemble learning. Furthermore, the R-squared values for RVM, BGRU, RF, NCV, and stacked ensemble learning were significantly higher at 15.456, 14.389, 11.809, and 6.591, respectively, compared to the R-CNN-TPOT's R-squared value of 0.095. The root mean squared error (RMSE) values for these techniques were also higher, with RVM at 3.406, BGRU at 2.751, RF at 2.162, NCV at 1.219, and the R-CNN-TPOT achieving the lowest RMSE of 0.251. The application of diverse features facilitated the identification of key representations within the EEG signals, and the hybrid ML approach of R-CNN-TPOT effectively captured the non-linear characteristics of the EEG data, resulting in reduced prediction errors.

Table 1. Comparative discussion

Variations	***Metrics***	***Methods***				
		RVM	***BGRU***	***RF***	***NCV and stacked ensemble learning***	***Proposed R-CNN-TPOT***
Training set	***MAE***	12.303	10.311	6.124	4.320	0.043
	MSE	11.602	7.566	4.674	1.485	**0.063**
	R-squared	15.098	14.322	13.972	8.597	0.099
	RMSE	3.406	2.751	2.162	1.219	**0.251**
K-value	***MAE***	11.496	10.005	7.141	4.920	**0.033**
	MSE	10.494	8.868	5.991	1.886	0.075
	R-squared	15.456	14.389	11.809	6.591	**0.095**
	RMSE	3.239	2.978	2.448	1.373	0.274

5. CONCLUSION

This study introduces an innovative method for estimating brain age from EEG signals utilizing TPOT. The prediction process involves extracting various features, including spectral, statistical, frequency domain, and decomposition domain characteristics from the EEG data. To enhance the quality of the EEG signal, a Gaussian filter is applied during the pre-processing phase to remove artifacts. Additionally, a feature selection technique based on Canberra distance is employed to reduce the dimensionality of the generated feature set, allowing for the identification of significant features within the EEG signal. Brain age estimation is performed using the R-CNN-TPOT, a hybrid machine learning model that integrates the mathematical framework of a CNN with TPOT through regression modeling based on fully connected layers. The effectiveness of the R-CNN-TPOT in predicting brain age is evaluated using metrics such as MAE, MSE, R-Squared, and RMSE, with experimental results indicating MAE at 0.033, MSE at 0.063, R-Squared at 0.095, and RMSE at 0.251. Future research directions include the exploration of hybrid deep learning techniques to enhance performance further.

REFERENCES

[1] Engemann, D.A., Mellot, A., Höchenberger, R., Banville, H., Sabbagh, D., Gemein, L., Ball, T. and Gramfort, A., "A reusable benchmark of brain-age prediction from M/EEG resting-state signals", Neuroimage, vol.262, pp.119521, 2022.

[2] Beheshti, Iman, M. A. Ganaie, Vardhan Paliwal, Aryan Rastogi, Imran Razzak, and Muhammad Tanveer. "Predicting brain age using machine learning algorithms: A comprehensive evaluation." *IEEE Journal of Biomedical and Health Informatics* 26, no. 4 (2021): 1432-1440.

[3] Madden, D.J., Jain, S., Monge, Z.A., Cook, A.D., Lee, A., Huang, H., Howard, C.M. and Cohen, J.R., "Influence of structural and functional brain connectivity on age-related differences in fluid cognition", Neurobiology of aging, vol.96, pp.205-222, 2020.

[4] Baecker, Lea, Jessica Dafflon, Pedro F. Da Costa, Rafael Garcia-Dias, Sandra Vieira, Cristina Scarpazza, Vince D. Calhoun, Joao R. Sato, Andrea Mechelli, and Walter HL Pinaya. "Brain age prediction: A comparison between machine learning models using region-and voxel-based morphometric data." *Human brain mapping* 42, no. 8 (2021): 2332-2346.

[5] Lu, A.T., Hannon, E., Levine, M.E., Crimmins, E.M., Lunnon, K., Mill, J., Geschwind, D.H. and Horvath, S., "Genetic architecture of epigenetic and neuronal ageing rates in human brain regions", Nature communications, vol.8, no.1, pp.15353, 2017.

[6] Al Zoubi, Obada, Chung Ki Wong, Rayus T. Kuplicki, Hung-wen Yeh, Ahmad Mayeli, Hazem Refai, Martin Paulus, and Jerzy Bodurka. "Predicting age from brain EEG signals—a machine learning approach." *Frontiers in aging neuroscience* 10 (2018): 184.

[7] Sun, H., Paixao, L., Oliva, J.T., Goparaju, B., Carvalho, D.Z., van Leeuwen, K.G., Akeju, O., Thomas, R.J., Cash, S.S., Bianchi, M.T. and Westover, M.B., "Brain age from the electroencephalogram of sleep", Neurobiology of aging, vol.74, pp.112-120, 2019.

[8] Liem, F., Varoquaux, G., Kynast, J., Beyer, F., Masouleh, S.K., Huntenburg, J.M., Lampe, L., Rahim, M., Abraham, A., Craddock, R.C. and Riedel-Heller, S., "Predicting brain-age from multimodal imaging data captures cognitive impairment", Neuroimage, vol.148, pp.179-188, 2017.

[9] Jeong, Hyerin, Ukeob Park, and Seung Wan Kang. "Region-specific Brain Age Prediction Models for Children and Adolescents Derived by Machine Learning." (2021). Tanveer, Muhammad, M. A. Ganaie, Iman Beheshti, Tripti Goel, Nehal Ahmad, Kuan-Ting Lai, Kaizhu Huang, Yu-Dong Zhang, Javier Del Ser, and Chin-Teng Lin. "Deep learning for brain age estimation: A systematic review." *Information Fusion* 96 (2023): 130-143.

[10] Baecker, L., Garcia-Dias, R., Vieira, S., Scarpazza, C. and Mechelli, A., "Machine learning for brain age prediction: Introduction to methods and clinical applications", EBioMedicine, vol.72, pp.103600, 2021.

[11] Olson, R.S. and Moore, J.H., "TPOT: A tree-based pipeline optimization tool for automating machine learning", In Workshop on automatic machine learning, PMLR, pp. 66-74, December 2016.

[12] Cook, Zara, Grant Sinha, Jack Wang, Chengzong Zhao, Nabil Belacel, Sam Doesburg, George Medvedev, UrsRibary, Vasily Vakorin, and Pengcheng Xi. "Enhancing Brain Age Prediction: A Generative AI Approach for EEG Machine Learning Models." In *2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1-6. IEEE, 2024.

[13] Lee, J.G., Kim, B.K., Jang, S.B., Yeon, S.H. and Ko, Y.W., "Accuracy enhancement of RSSI-based distance estimation by applying Gaussian filter", Indian Journal of Science and Technology, vol.9, no.20, pp.1-5, 2016.

[14] Anemüller, J., Schmidt, D. and Bach, J.H., "Detection of speech embedded in real acoustic background based on amplitude modulation spectrogram features", In Ninth Annual Conference of the International Speech Communication Association, 2008.

[15] Hassan, A.R., Siuly, S. and Zhang, Y., "Epileptic seizure detection in EEG signals using tunable-Q factor wavelet transform and bootstrap aggregating", Computer methods and programs in biomedicine, vol.137, pp.247-259, 2016.

[16] Hong, K.S., Khan, M.J. and Hong, M.J., "Feature extraction and classification methods for hybrid fNIRS-EEG brain-computer interfaces", Frontiers in human neuroscience, vol.12, pp.246, 2018.

[17] Diykh, M., Li, Y. and Wen, P., "EEG sleep stages classification based on time domain features and structural graph similarity", IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol.24, no.11, pp.1159-1168, 2016.

[18] Singh, A.K. and Krishnan, S., "Trends in EEG signal feature extraction applications", Frontiers in Artificial Intelligence, vol.5, 2022.

[19] Androutsos, D., Plataniotiss, K.N. and Venetsanopoulos, A.N., "Distance measures for color image retrieval", In Proceedings IEEE International Conference on Image Processing. ICIP98 (Cat. No. 98CB36269), vol. 2, pp. 770-774, October 1998.

[20] Nguyen, H.T., Li, S. and Cheah, C.C., "A layer-wise theoretical framework for deep learning of convolutional neural networks", IEEE Access, vol.10, pp.14270-14287, 2022.

[21] Mannepalli, K., Sastry, P.N. and Suman, M., "A novel adaptive fractional deep belief networks for speaker emotion recognition", Alexandria Engineering Journal, vol.56, no.4, pp.485-497, 2017.

[22] Murugappan, M. and Murugappan, S., "Human emotion recognition through short time Electroencephalogram (EEG) signals using Fast Fourier Transform (FFT)", In Proceedings of 2013 IEEE 9th International Colloquium on Signal Processing and its Applications, pp. 289-294, IEEE, March 2013.

[23] Ji, N., Ma, L., Dong, H. and Zhang, X., "EEG signals feature extraction based on DWT and EMD combined with approximate entropy", Brain sciences, vol.9, no.8, pp.201, 2019.

[24] Trabelsi, A., Elouedi, Z. and Lefevre, E., "Decision tree classifiers for evidential attribute values and class labels", Fuzzy Sets and Systems, vol.366, pp.46-62, 2019.

[25] Bhaladhare, P.R. and Jinwala, D.C., "A clustering approach for the-diversity model in privacy preserving data mining using fractional calculus-bacterial foraging optimization algorithm", Advances in Computer Engineering, 2014.

[26] HarMNqEEG dataset is obtained from "synapse.org", accessed on June 2022.

[27] Li, M., Wang, Y., Lopez-Naranjo, C., Hu, S., Reyes, R.C.G., Paz-Linares, D., Areces-Gonzalez, A., Abd Hamid, A.I., Evans, A.C., Savostyanov, A.N. and Calzada-Reyes, A., "Harmonized-Multinational qEEG norms (HarMNqEEG)", NeuroImage, vol.256, pp.119190, 2022.