Birla Institute of Science and Technology

Deep Neural Network

Assignment-1

Submitted by Group-111

2023ac05875	Monica Malik	100%
2023ac05912	Saurabh Jalendra	100%
2023ac05573	Tushar Shandilya	100%
2023ac05862	Reddy Balaji C	100%

PART-A

Literature Exploration and Comparison

EEG based Brain Machine Interfaces

Title of the paper	A Fuzzy Ensemble-Based Deep Learning Model for EEG-Based Emotion	MSHANet: A Multi-Scale Residual Network with Hybrid Attention for Motor Imagery EEG Decoding	An End-to-End Brain Computer Interface System for Mental Workload Estimation through Hybrid Deep Learning Model
Authors	Trishita Dhara, Pawan Kumar Singh, Mufti Mahmud	Mengfan Li, Jundi Li, Xiao Zheng, Jiahao Ge, and Guizhi Xu.	Vipul Sharma, Mitul Kumar Ahirwal
Year of publication	2023	2024	2024
Architecture of Deep Learning (including the number of layers, types of layers, activation functions, and any unique features)	First candidate model is a hybrid of **CNN and LSTM** models. The **CNN layers extract spatial features** from the signals, and the **LSTM part extracts temporal features.** Fully connected layers follow the LSTM layers, and then the final prediction is obtained. **Hybrid CNN-LSTM:** Extracts spatial features with CNN and temporal features with LSTM, followed by dense layers. **Layers:** **Hybrid CNN-GRU:** Similar to CNN-LSTM but uses GRU for reduced training time.	The MSHANet architecture comprises a **convolution block, sliding window, hybrid attention block, and multi-scale residual block**. EEG data is processed through the convolution block with **temporal and spatial filters** to extract features into a compact convolutional network, which is fed into a **multi-branch network.** Each branch employs a **hybrid attention block combining multi-head and SE attention mechanisms** to highlight important features. A multi-scale residual block with **temporal convolutional networks (TCNs)** of varying kernel sizes extracts deeper	**Layers**: The architecture consists of two 1D-CNN layers followed by two BLSTM layers, a dense layer, and an output layer. **1D-CNN Layers**: Extract spatial features; the first layer has 32 filters (kernel size = 16 and stride length of 1), and the second has 16 filters (kernel size = 8 and stride length of 1), both with **ReLU activation**. **BLSTM Layers**: The output of the stacked 1D CNN layers is then passed to BLSTM layers. The first and second BLSTM layer has 32 neurons with a tan(h) activation function each. **Dense Layers**: The second BLSTM layer generates a single vector as its output which is fed to a dense

	1D-CNN: Focused on feature extraction with sequential dense layers for classification **Activation functions:** ReLU and Sigmoid. **Dropout layers and max pooling** are used for regularization and down-sampling. **Final model**: **Fuzzy ensemble using Gompertz function** to combine predictions adaptively	information, while a separate TCN shares parameters across branches to capture common features in different time windows. The fused high-dimensional features are classified using a **softmax layer**. **Layers:** **Convolutional Block:** Two temporal filters and one spatial filter with ReLU activation. Dropout is used for regularization. **Hybrid Attention Block:** Combines Multi-Head Attention (MHA) and Squeeze-and-Excitation (SE) attention for refined feature extraction. **Multi-Scale Residual Block:** Temporal Convolutional Networks (TCN) with different kernel sizes to capture both local and global features. Shared TCN parameters reduce network complexity	layer for classification. The final dense layer has 40 neurons with ReLU activation, and the output layer uses Softmax for classification into binary or ternary classes. **Unique Features**: The end-to-end architecture avoids handcrafted feature engineering, focusing entirely on automated feature extraction
How is the network helping the overall task? e.g., feature engineering or classification or regression or all	The proposed ensemble method generates fuzzy ranks of the different models using the Gompertz function. It fuses the decision scores adaptively from those models to make the combined prediction on the test set. weightage is assigned dynamically to each model based on the confidence measure. For each model, indivisual confidence (c) is found, which is normalised to get a normalised confidence value. These confidence scores are used for calculating the fuzzy rank using the Gompertz function. Let Mj denote the top M ranks for a particular class c . If the rank does not belong to the top M ranks,	**Convolutional block** Two temporal filters and one spatial filter are used in this block. The temporal filters are used for 2 D convolution and batch normalization. The spatial filter is used to reduce the spatial dimensions of the inputs, and a regularization technique dropout is added to prevent overfitting. **Sliding Window** This is mainly used to separate the feature into sub-features which serve as inputs for different branches. **Hybrid Attention Block** Multi-head attention mechanism (**MHA**) is used to learn different features in the	-The network performs feature extraction and classification for mental workload estimation. -It classifies mental workload into binary (low vs. high) and ternary (low, moderate, high) categories, achieving high accuracy. A deep learning model for the multivariate time series i.e., EEG signals, classification into 3 and 2 classes has been developed. The model consists of **1D convolution (1DCNN) layers** followed by **bidirectional long short-term memory (BLSTM) layers** for feature extraction. A fully connected neural network to the output of these layers is also used for classification. The use of deep learning has allowed for the classification of complex multichannel EEG data without the need for handcrafted feature

	then two penalty values are calculated, P1 and P2. Next, we calculate two more factors, the rank-sum (Sj) and the complement of confidence score factor (Fj). The final score (SCj) is the product of Sj and Fj SCj =Sj×Fj Finally, the resultant class (c) is the class with a mini mum SCj value which gives the final decision score of the proposed ensemble model. The Model Assists in feature extraction, classification of emotions based on valence and arousal Improves prediction accuracy with subject-dependent and independent approaches. Provides robust performance across benchmark datasets	input sequence in parallel. The input features are first processed with multiple sets of self-attention, and then the k-head results are stitched together and linearly transformed to obtain the final output. Squeeze and Excitation (**SE**) attention is used to further enhance the feature representation and selection and remove the redundant features present after the MHA. **Multi-scale residual block** This extracts feature-rich and robust features from EEG through convolution kernels at different scales. This consists of temporal convolutional networks (TCN) with different convolutional kernel sizes further extract deeper and richer information. **Soft Max Layer** Shared TCN is used to share network parameters across multiple branches to extract common features in different time windows and these high dimensional features are given as input to the SoftMax layer for classification.	extraction, demonstrating the power of deep learning. The **CNN-BLSTM model** is used in the experiment for both binary and ternary **classification of EEG signals**, this model learns both the spatial and the temporal characteristics of multichannel EEG signals to do automated feature extraction.
Training procedures (e.g, training strategy, including optimization algorithms, learning rates, batch sizes, and regularization techniques)	This paper proposes an **ensemble learning approach for emotion detection from EEG signals**. We have trained three individual models and combined them **using the fuzzy ensemble technique and max voting ensemble.** Each model has been trained individually for **50 epochs** for both datasets before combining them. **Optimization Algorithm**: Adam optimizer with a **learning rate of 0.001** **Training strategy:** Subject-	**Training** The trials from the two sessions of each subject are mixed and 288 are randomly taken as the training set and the remaining 288 as the test set. All trials are standardized to the function: *X =(E−mean(E))/std(E)* *i= 1,2,…,C* C - denotes channel number T - denotes time point number $E_j \in R^{C\times T}$	**Optimization Algorithm**: Stochastic Gradient Descent (SGD) with cross-entropy loss **Learning rate** is chosen in a specific way as described by Leslie N. Smith(Cyclical learning rates for training neural networks). Initial training started with an **initial learning rate of 1e-7 and exponentially increased in each epoch using the formulae** Learning rate =**10**7+(epoch/40)**-Binary Classification, Learning rate =**10**7+(epoch/100)**-ternary Classification **Batch Size**: Selected as factors of

	independent and subject-dependent approaches. **Batch sizes and training split: 60% training, 20% validation, and 20% testing** (For subject independent approach) For the subject-dependent approach, the model is trained and tested (**60:20:20 ratio**) separately for **each subject**, and then the average of the results was taken. **Regularization:** Dropout layers in convolutional and fully connected layers	It uses **Adam as the optimizer**, the **batch size** is set to **64**, the **learning rate** is set to **5e-4**, and the number of iterations is fine-tuned with the different methods. **Cross-Validation:** Five-fold cross-validation was performed to validate model performance and ensure robustness across diverse training subsets. **Regularization:** Dropout layers were included within the convolution block to prevent overfitting and enhance generalization. Parameter sharing in the multi-scale residual block reduced network complexity, effectively minimizing overfitting risks.	the training data size for efficiency. Appropriate **batch size is selected from the factors of the size of training data** i.e., a number that evenly divides the training set. **Cross-Validation**: Models were trained with five-fold and seven-fold cross-validation to ensure robustness. **Regularization**: Dropout layers are included to prevent overfitting The resulting models have 51,370 and 51,411 trainable parameters respectively for **binary and ternary classification** tasks.
Evaluation / Performance metric used	**Metrics used:** **- Accuracy** and **F1-score**. - Results show accuracy of 98.73% and 98.39% for valence and arousal on the AMIGOS dataset (subject-independent), and 99.38% and 98.66% (subject-dependent). For the DEAP dataset, accuracy was 90.84% and 91.72% (subject-independent), and 95.78% and 95.97% (subject-dependent)	- **Accuracy:** Evaluates the classification performance. - **Cohen's Kappa:** Measures agreement for multiclass classification beyond chance. **-ROC-AUC:** Assesses classification performance for each motor imagery category. - **Standard Deviation (Std):** Indicates robustness across subjects - **Session-Specific Tasks:** MSHANet is compared with baseline models like EEGNet, EEGTCNet, ATCNet, and DeepConvNet. MSHANet achieves the highest average accuracy (83.18%) and kappa (0.78) for nine subjects in session-specific tasks. - **Cross-Session Tasks:** The average accuracy for MSHANet is 80.09%,	- **Accuracy=** (TP+TN)/ Total instances - **Precision=** TP /(TP+FP) - **Recall = =** TP /(TP+FN) - **F1-Score**- 2∗ (precision ∗ recall)/ (precision + recall) These evaluation parameters have been used for both Binary and ternary classification models - **Binary classification** achieved 97.89% accuracy (holdout), and ternary classification achieved 95.87% accuracy (holdout). - **Cross-validation accuracies:** 96.77% (binary, seven-fold CV) and 95.36% (ternary, seven-fold CV)

		surpassing baseline models such as ATCNet and TCNet-Fusion	
Name of Dataset used. If a public dataset, provide the URL.	**-DEAP Dataset**: EEG recordings of 32 subjects while watching music videos. Preprocessed data available in Python and MATLAB formats. URL: DEAP Dataset **- AMIGOS Dataset**: EEG, ECG, and GSR recordings from 40 participants in individual and group settings. URL: AMIGOS	**- Name:** BCI Competition IV 2a Dataset. **- Description:** EEG data from 9 subjects performing four motor imagery tasks (left hand, right hand, feet, and tongue). **URL:** DatasetLink	In this study, Simultaneous task **EEG workload (STEW) dataset** is used for the mental workload classification task. **STEW measures the mental workload during "no task" and the workload induced by "simultaneous capacity (SIMKAP)-based multi-tasking activity**". In STEW dataset 45 subjects' EEG recordings with their mental workload ratings during SIMKAP is provided. These ratings were binned into 2 and 3 classes respectively for binary and ternary classification. Dataset Links- https://doi.org/10.1109/TNSRE.2018.2872924 https://doi.org/10.21227/44r8-ya50
Conclusion	- The fuzzy ensemble model achieves state-of-the-art results on emotion recognition from EEG signals for both subject-independent and subject-dependent setups. - Future directions include developing multi-modal models and using image representations of EEG data to enhance efficiency	- The MSHANet model effectively decodes motor imagery EEG signals with robust performance, achieving high classification accuracy and stability across sessions. The hybrid attention block and multi-scale residual block enhance feature extraction and reduce network complexity. - Future research may focus on lightweight models and validating the approach on diverse datasets to improve generalization	- The proposed CNN-BLSTM model achieved state-of-the-art performance in mental workload classification, demonstrating the effectiveness of end-to-end deep learning for EEG-based tasks. - The lightweight model, with only ~50,000 parameters, ensures faster training and real-time applicability. - Future work will focus on testing diverse datasets and extending the model to classify more workload levels

Conclusion

After thoroughly analyzing and comparing the three selected research papers on **EEG-based Brain-Machine Interfaces (BMIs),** each paper presents valuable contributions to the field, especially in the context of deep learning models applied to **motor imagery, emotion recognition, and mental workload estimation**. Below is a summary of the findings, **highlighting the strengths and limitations of each model**.

Paper 1: ***MSHANet: A Multi-Scale Residual Network with Hybrid Attention for Motor Imagery EEG Decoding***

- **Strengths:** MSHANet employs multi-scale residual networks and hybrid attention mechanisms, providing robust performance in cross-session scenarios, which is essential for real-world applications in periprosthetic and rehabilitation. Its high classification accuracy (83.18% session-specific accuracy) in motor imagery decoding makes it a strong candidate for assistive technology, where variability across sessions and subjects is a common challenge.
- **Limitations:** The model's computational complexity and resource-intensive training requirements may limit its applicability in real-time systems or on devices with limited processing power.

Paper 2: ***A Fuzzy Ensemble-Based Deep Learning Model for EEG-Based Emotion Recognition***

- **Strengths:** The fuzzy ensemble-based approach enhances the robustness of emotion recognition from EEG signals, making the system more resilient to noise and variability. The use of multiple deep learning models combined in a fuzzy ensemble increases the accuracy and reliability of emotion classification, making it applicable for adaptive human-computer interaction (HCI) systems, personalized therapy, and affective computing.
- **Limitations:** The ensemble approach adds complexity, which might make real-time implementation challenging. Additionally, issues with cross-subject generalization and the need for large datasets still limit its scalability.

Paper 3: ***An End-to-End Brain-Computer Interface System for Mental Workload Estimation through Hybrid Deep Learning Model***

- **Strengths:** This paper presents a hybrid deep learning model combining 1D-CNN and BiLSTM for mental workload estimation from EEG signals. The model achieves high accuracy (96.77% for binary classification), showcasing its potential in applications such as driver safety systems and adaptive learning environments.
- **Limitations:** The reliance on manual feature extraction in some parts of the system may limit the model's flexibility in handling more complex and dynamic real-world environments, where workload fluctuates.

Key Insights and Future Research Directions:

- **Real-World Applicability:** MSHANet stands out as the most practical model for real-time applications in periprosthetic and rehabilitation. The model's ability to perform well across different sessions, despite variations in EEG data, makes it ideal for assistive technologies for individuals with motor disabilities. However, improving its computational efficiency for real-time applications is a key challenge to address in future work.

- **Trade-offs Between Models:** While MSHANet provides excellent accuracy, its complexity may pose challenges in mobile or wearable devices. On the other hand, the Fuzzy Ensemble Model enhances robustness but sacrifices real-time processing efficiency due to its ensemble configuration. Paper 3 shows promising results for mental workload estimation, but further improvements are needed to enhance its generalization and flexibility in multi-task scenarios.

- **Dataset Limitations:** A common limitation across all three models is the reliance on specific datasets, which may not fully represent diverse populations. Expanding the datasets to include a broader range of participants and environments will improve the generalizability of these models.

Future Work:

- For MSHANet, future research could focus on optimizing the model for real-time performance without compromising accuracy.
- For emotion recognition, research could explore methods to simplify ensemble models while maintaining performance and enhancing cross-subject generalization.
- For mental workload estimation, improving multi-task adaptability and addressing feature extraction limitations will be important next steps.

Final Thoughts:

In conclusion, all three papers make significant contributions to the field of EEG-based Brain-Machine Interfaces (BMIs). MSHANet offers the most promising approach for motor imagery decoding and periprosthetic, but further work is needed to enhance its real-time scalability. Emotion recognition and mental workload estimation also hold great potential for improving adaptive systems and human-computer interaction. By focusing on model efficiency, generalization, and dataset diversity, future research can push the boundaries of EEG-based BCIs and enable practical, real-world applications.