1. RESULTS AND DISCUSSION

The three-dimensional tensor containing the data of experiments is split randomly into training and testing sets where 75% of data is used for training and 25% is used for testing. The testing dataset is a 3D tensor with the shape of (number of testing samples, number of EEG channels, time) each testing sample has its corresponding label. Before sending the data into the CNN for testing, pre-processing steps like standard scaling is done as it was done for the training dataset.

Testing data is then passed through the model to generate the prediction labels, these predictions are compared with the true labels to evaluate the model, and evaluation parameters such as accuracy, recall, and precision are obtained from these labels. Each subject is evaluated independently, and the models' accuracies are obtained for every subject separately.

To evaluate the performance of the proposed algorithms in handling the non-stationarity in intersession data we have used dataset II a from the BCI competition IV.
The dataset consists of EEG signals recorded from nine subjects using 22 electrodes. For these datasets, the raw EEG data is extracted from 0s to 4s, sampled at 250Hz. The motor-imagery tasks performed by the subjects are divided into four classes, namely, left hand (class 1), right hand (class 2), foot (class 3), and tongue (class 4). Now we consider two classes at a time and classify them. The binary classifications are left versus right-hand classification, left hand versus foot classification, left hand versus tongue classification, right hand versus foot classification, right hand versus tongue classification, and foot versus tongue classification. In all these binary classification experiments, each classification has 24 trials for one subject. Considering 144 trails (6 classifications) for each subject we have a total of 1296 trails in the training and testing data.

PERFORMANCE EVALUATION

The proposed algorithms are EEG Network, Deep CNN Network, and Shallow Convolution Network. The performance of the proposed algorithms EEG Network, Deep CNN Network, and Shallow Convolution Network is compared with the RoCSP-SRIT2NFIS algorithm.

1.1. EEG NETWORK

The performance of the EEG Network for the six classification tasks is presented in Table 1. From Table 1 we can see the SM (in %) for the six classification tasks, left versus right hand, left hand versus foot, left hand versus tongue, right hand versus foot, right hand versus tongue, and foot versus tongue are 88.8, 89.2, 90.2, 88.8, 89.2, and 88.8 respectively.

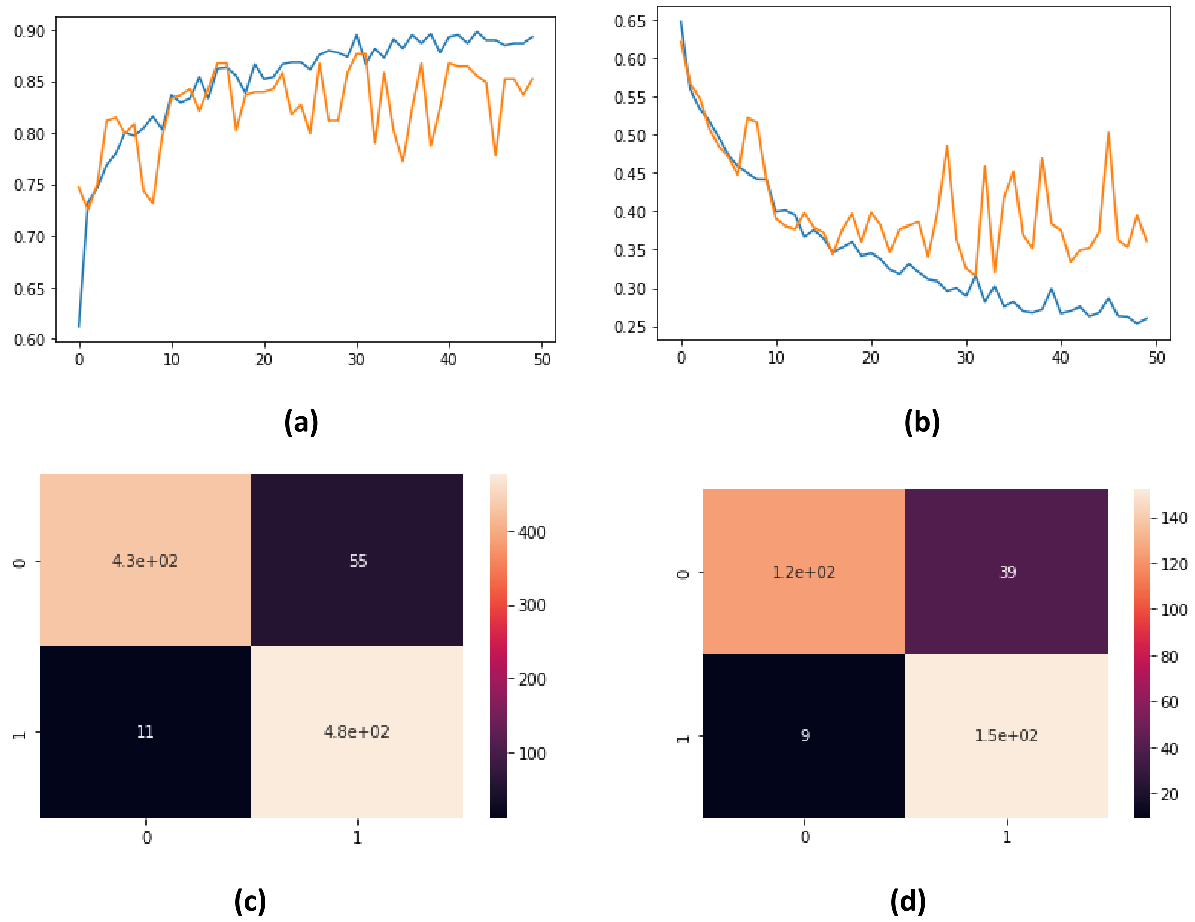

Fig. 1. Accuracy, Loss, and Confusion matrices for training and testing for right vs foot classification task of EEG Network. The blue curve represents the training curve and the orange line represents the testing curve with the x-axis representing the no. of epochs and the y-axis representing the accuracy and loss (in %). (a) Accuracy graph for training and testing. (b) Loss graph for training and testing. (c) Confusion Matrix for testing data. (d) Confusion Matrix for testing data.

From Fig. 1. We can see the graphs of accuracy and loss for the EEG Network, we got the highest Validation Accuracy of 88.27% at the 31st epoch. EEG Network achieved the highest accuracy of 90.2% for Left versus tongue classification. From Table II we can see that the overall average Accuracy (in %) of the EEG Network is 89.16.

1.2. SHALLOW CONVOLUTION NETWORK

The performance of the Shallow Convolution Network for the six classification tasks is presented in Table 1. From Table 1 we can see the SM (in %) for the six classification tasks, left versus right hand, left hand versus foot, left hand versus tongue, right hand versus foot, right hand versus tongue, and foot versus tongue are 88.46, 89.23, 90.0, 89.2, 88.8, and 88.07 respectively.

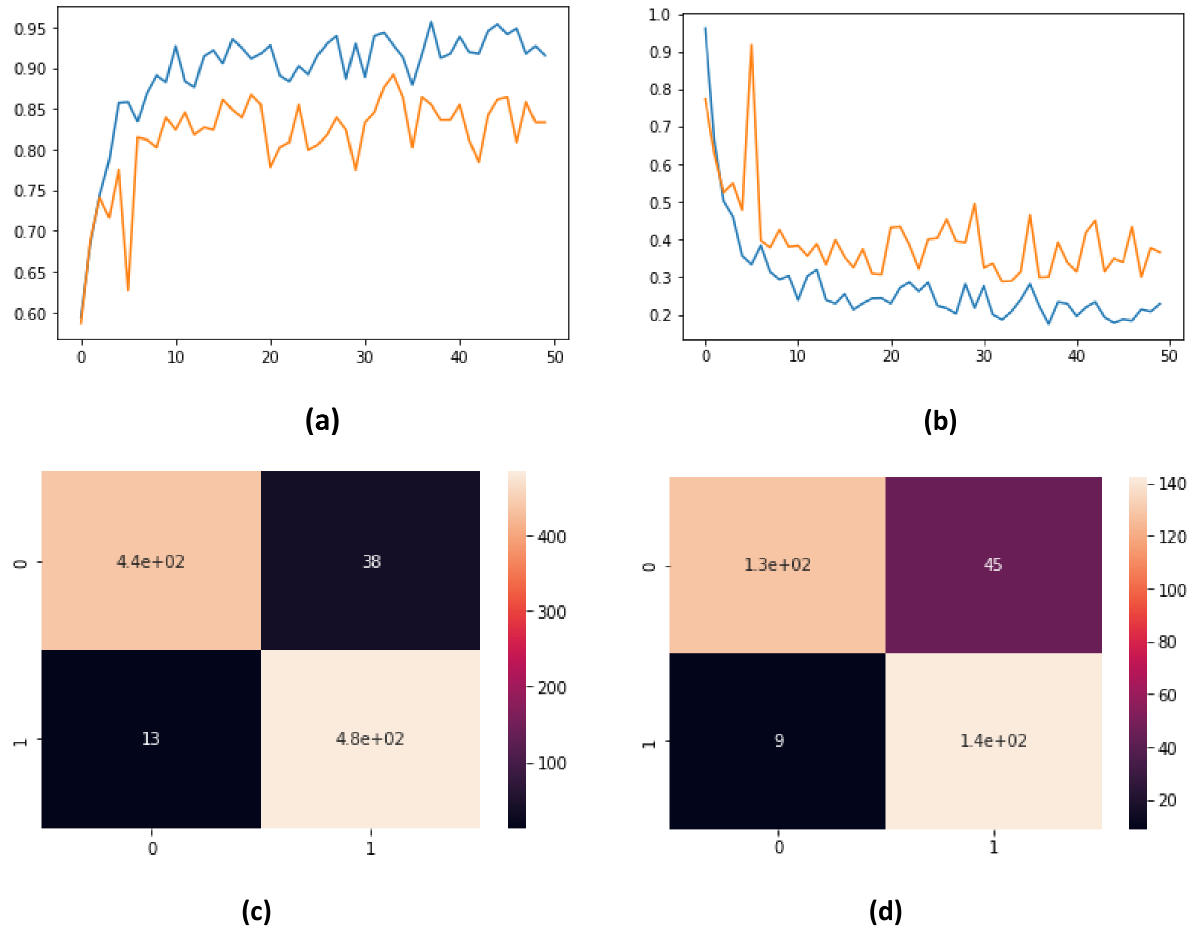

Fig. 2. Accuracy, Loss, and Confusion matrices for training and testing for right vs foot classification task of Shallow Convolution Network. The blue curve represents the training curve and the orange line represents the testing curve with the x-axis representing the no. of epochs and the y-axis representing the accuracy and loss (in %). (a) Accuracy graph for training and testing. (b) Loss graph for training and testing. (c) Confusion Matrix for testing data. (d) Confusion Matrix for testing data.

From Fig. 2. We can see the graphs of accuracy and loss for the Shallow Convolution Network, we got the highest Validation Accuracy of 89.23% at the 34th epoch. Shallow Convolution Network achieved the highest accuracy of 90.0% for Left versus tongue classification. From Table II we can see that the overall average Accuracy (in %) of the Shallow Convolution Network is 88.96%.

1.3. DEEP CNN NETWORK

The performance of the Deep CNN Network for the six classification tasks is presented in Table 1. From Table 1 we can see the SM (in %) for the six classification tasks, left versus right hand, left hand versus foot, left hand versus tongue, right hand versus foot, right hand versus tongue, and foot versus tongue are 88.46, 88.04, 91.92, 90.31, 88.52, and 88.89 respectively.

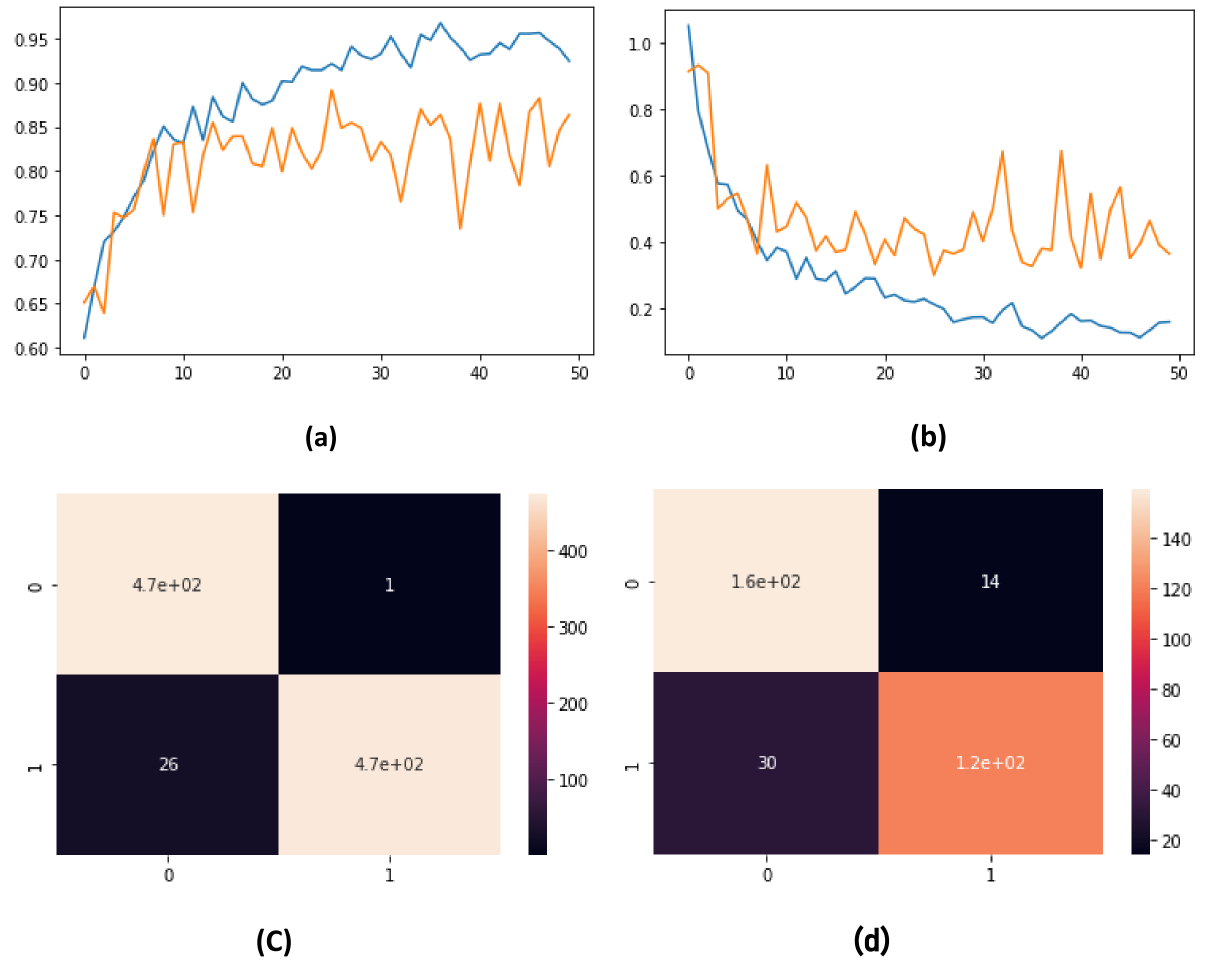

Fig. 3. Accuracy, Loss, and Confusion matrices for training and testing for right vs foot classification task of Deep CNN Network. The blue curve represents the training curve and the orange line represents the testing curve with the x-axis representing the no. of epochs and the y-axis representing the accuracy and loss (in %). (a) Accuracy graph for training and testing. (b) Loss graph for training and testing. (c) Confusion Matrix for testing data. (d) Confusion Matrix for testing data.

From Fig. 2. We can see the graphs of accuracy and loss for the Deep CNN Network, we got the highest Validation Accuracy of 88.89% at the 26th epoch. Deep CNN Network achieved the highest accuracy of 91.92% for Left versus tongue classification. From Table II we can see that the overall average Accuracy (in %) of the Deep CNN Network is 89.35%.

COMPARISON

The proposed algorithms are compared with the RoCSP-SRT2NFIS algorithm over multiple datasets. We have a total of 54 binary classification experiments with nine subjects. We used the subject mean (SM in %) to compare the algorithms with the RoCSP-SRT2NFIS algorithm. From Table 1 we can see that all our proposed algorithms are outperforming the RoCSP-SRT2NFIS algorithm. It should be noted that by this our algorithms are proved to be more robust towards the artifacts and non-stationarities in the EEG data.

TABLE I

PERFORMANCE OF THE ALGORITHMS FOR THE SIX CLASSIFICATION TASKS ON THE BCI COMPETITION IV DATASET IIA

Subjects	(a) left versus right			
	RoCSP-SRIT2NFIS	EEG NETWORK	DEEP CNN NETWORK	SHALLOW CONV NETWORK
A1	93.06	85.7	82.1	92.9
A2	68.75	71.9	87.5	65.6
A3	97.22	88.9	92.6	92.6
A4	75.00	79.3	89.7	82.8
A5	65.97	100	92.6	92.6
A6	72.22	88.6	82.9	85.7
A7	86.11	96.2	96.2	92.3
A8	97.22	96.9	84.4	96.9
A9	93.75	91.7	91.7	100
SM	83.26	88.8	88.46	88.46

Subjects	(b) left hand versus foot			
	RoCSP-SRIT2NFIS	EEG NETWORK	DEEP CNN NETWORK	SHALLOW CONV NETWORK
A1	99.31	96.4	89.3	96.4

A2	83.33	78.1	87.5	81.3
A3	95.83	88.9	85.2	92.6
A4	90.28	86.2	79.3	86.2
A5	70.83	96.3	88.9	100
A6	72.22	85.7	82.9	82.9
A7	98.61	96.2	100	100
A8	91.67	81.3	84.4	90.6
A9	97.92	100	100	75.0
SM	88.89	89.2	88.04	89.23

Subjects	(c) left hand versus Tongue			
	RoCSP-SRIT2NFIS	EEG NETWORK	DEEP CNN NETWORK	SHALLOW CONV NETWORK
A1	99.31	92.9	96.4	89.3
A2	73.61	93.8	84.4	81.3
A3	96.53	100	100	88.9
A4	92.36	100	100	96.6
A5	78.47	66.7	88.9	88.9
A6	75.00	77.1	77.1	82.9
A7	97.22	88.5	88.5	92.3
A8	96.53	93.8	96.9	93.8
A9	97.92	100	100	100
SM	89.66	90.2	91.92	90.0

Subjects	(d) right hand versus foot			
	RoCSP-SRIT2NFIS	EEG NETWORK	DEEP CNN NETWORK	SHALLOW CONV NETWORK
A1	100	92.8	92.8	78.5
A2	84.03	84.3	81.2	75.0
A3	95.83	81.4	88.8	96.2
A4	90.97	86.2	96.5	93.1
A5	72.92	77.7	92.5	96.2
A6	70.83	94.2	88.5	88.5
A7	99.30	96.1	88.4	92.3
A8	90.28	90.6	93.7	93.7
A9	89.58	95.8	91.6	91.6
SM	88.19	88.8	90.31	89.2

Subjects	(e) right hand versus Tongue			
	RoCSP-SRIT2NFIS	EEG NETWORK	DEEP CNN NETWORK	SHALLOW CONV NETWORK
A1	100	92.8	93.0	92.8
A2	72.92	93.7	80.1	87.5

A3	99.30	92.5	83.7	85.1
A4	88.89	86.2	88.9	86.2
A5	77.08	88.8	93.1	92.5
A6	74.31	94.2	87.9	80.0
A7	98.61	92.3	89.4	96.1
A8	91.67	71.8	90.4	90.6
A9	95.83	91.6	91.8	91.6
SM	88.73	89.2	88.52	88.8

Subjects	(f) foot versus Tongue			
	RoCSP-SRIT2NFIS	EEG NETWORK	DEEP CNN NETWORK	SHALLOW CONV NETWORK
A1	78.47	85.7	79.9	85.7
A2	79.86	90.6	82.5	87.5
A3	79.17	88.8	89.8	85.1
A4	73.61	89.6	92.6	89.6
A5	75.00	81.4	88.9	81.4
A6	75.00	91.4	89.6	91.4
A7	86.11	92.3	95.3	96.1
A8	90.28	84.3	89.5	87.5
A9	95.13	95.8	93.6	87.5
SM	81.40	88.8	88.89	88.07

Results from Table 1 show that the performance of our algorithms is more consistent when compared to that of the RoCSP-SRIT2NFIS algorithm on classification tasks like left versus right and foot versus tongue.

We can also observe that for a more complex binary classification task like foot vs tongue our algorithms are providing 6-7% better performance than the RoCSP-SRIT2NFIS.

TABLE II

COMPARISON OF THE TOTAL AVERAGE PERFORMANCE OF THE ALGORITHMS FOR THE SIX CLASSIFICATION TASKS ON THE BCI COMPETITION IV DATASET IIA

Algorithms	Total average Accuracy
RoCSP-SRIT2NFIS	86.68
EEG Network	89.16
Shallow Conv Network	88.96
Deep CNN Network	89.35

From Table 2 we can consider that all the proposed algorithms are performing better than the RoCSP-SRIT2NFIS algorithm as we can see that the total average (in %) of all our algorithms is greater than that of the RoCSP-SRIT2NFIS algorithm. We can also conclude that our proposed algorithms are performing 3-4 % better than the next best-performing algorithm.