King Saud University

College of Computer and Information Science

Information Technology

Bangalore EEG Epilepsy

Section:56547

Student	ID
Dana Sultan Alotaibi	445203123
Jana Saleh Obaid	445200286

5.1 Decision Tree Classification

5.1.1 Overview of Decision Trees:

A Decision Tree is a supervised classification model that predicts outcomes by splitting data based on feature values. In this project, the model was used to classify EEG signals, where each split reflects the value of an EEG channel. The most important features appear at the top of the tree because they reduce impurity the most.

Both the Gini and Entropy trees generated for our dataset consistently used Channel8, Channel11, and Channel7 as the first splitting features, indicating these channels are the most informative for distinguishing EEG classes. To improve interpretability, the tree depth was limited to four levels, producing readable visualizations for both criteria gini and entropy.

5.1.2 Gini Impurity:

Gini Impurity measures how mixed the classes are inside a node. A value of 0.0 means the node is perfectly pure, while higher values indicate more class variation. The Decision Tree chooses splits that reduce Gini impurity the most, creating purer child nodes at each step.

```
DecisionTreeClassifier(max_depth=4, criterion="gini", random_state=123)
```

- Channel8 was chosen as the root split, meaning it provided the greatest reduction in impurity.

- Channel7 and Channel11 appeared repeatedly in upper levels, confirming they are highly informative EEG features.

- Many leaf nodes reached Gini = 0.0, showing perfectly pure class predictions.

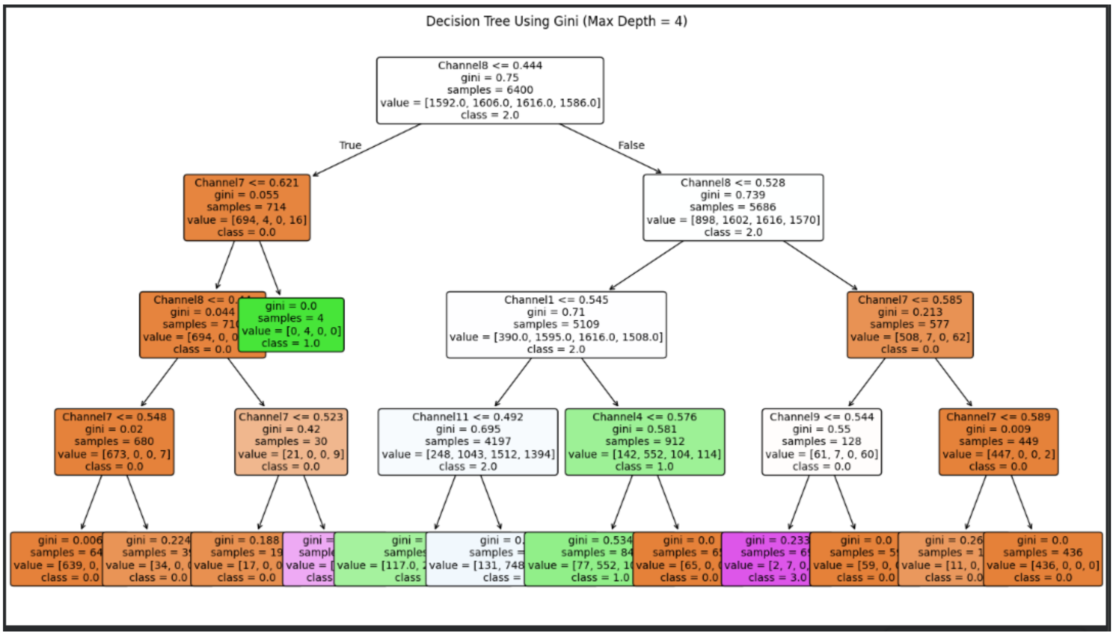
- Limiting the tree to max_depth = 4 kept the structure readable while still capturing the key decision patterns.

5.1.3 Entropy / Information Gain

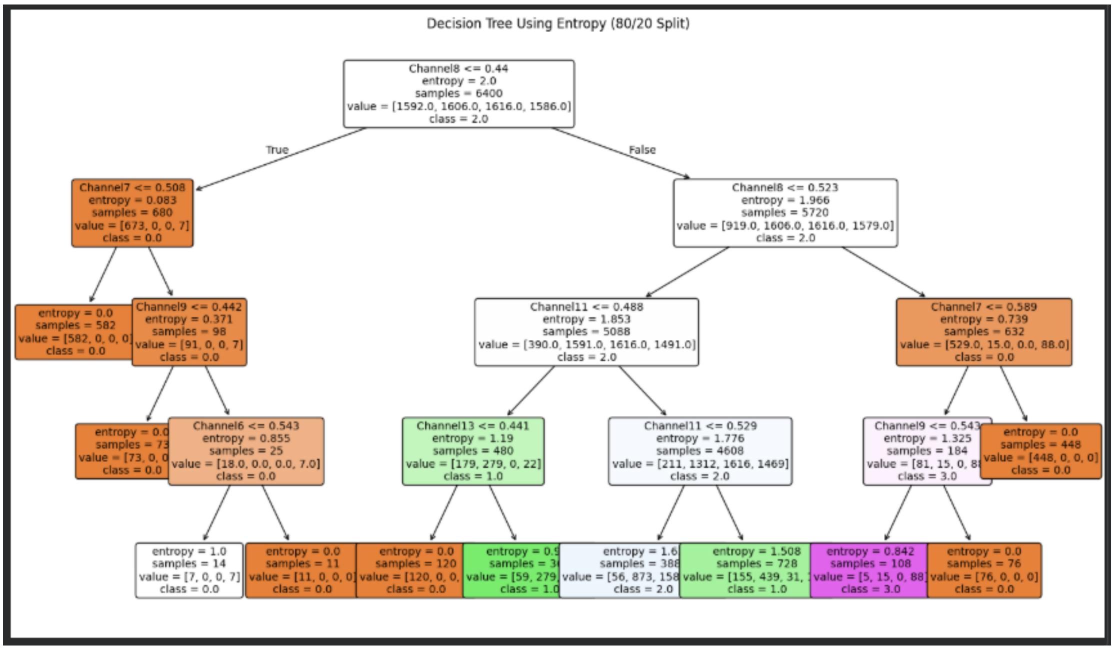
Entropy measures the disorder within a node, and the Decision Tree selects splits that give the highest Information Gain. In our entropy-based model (max_depth = 4), Channel8, Channel11, and Channel7 were the most influential features, appearing in the top splits and providing the greatest reduction in entropy. Several leaf nodes reached entropy = 0.0, indicating perfectly pure class predictions. The entropy model achieved strong performance across all splits, with its best accuracy (0.88125) at the 80/20 split.

5.1.4 Decision Tree Visualizations:

Decision Tree (Gini) for split 80/20



Decision Tree (Entropy) for split 80/20



6.1 Classification Evaluation

6.1.1: Accuracy Comparison Between Gini and Entropy Criteria:

Accuracy for gini:

	Train/Test Split	Accuracy (Gini)
0	90/10	0.875000
1	80/20	0.864375
2	70/30	0.849167

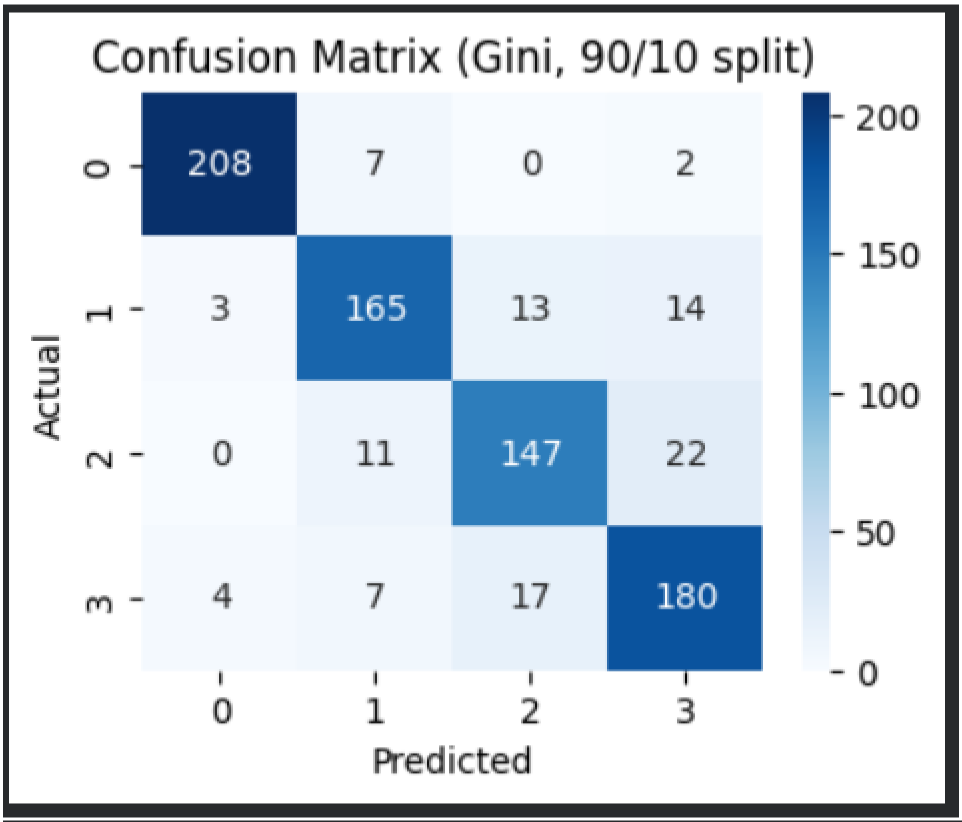
Accuracy for Entropy:

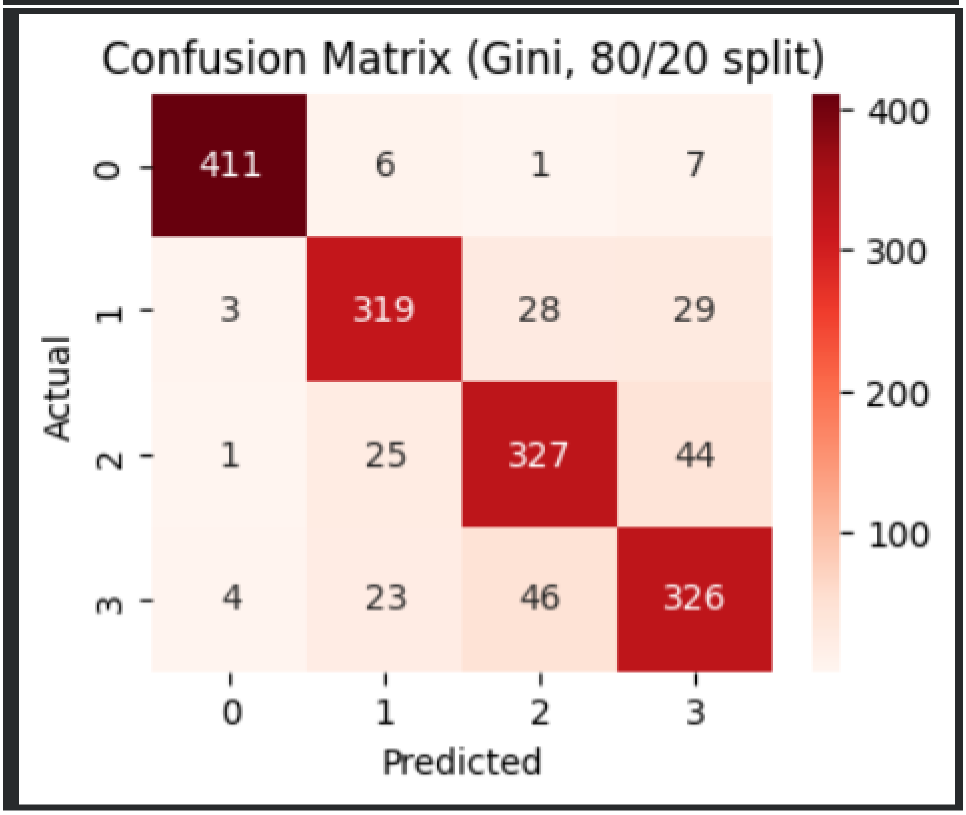
	Train/Test Split	Accuracy (Entropy)
0	90/10	0.873750
1	80/20	0.881250
2	70/30	0.857917

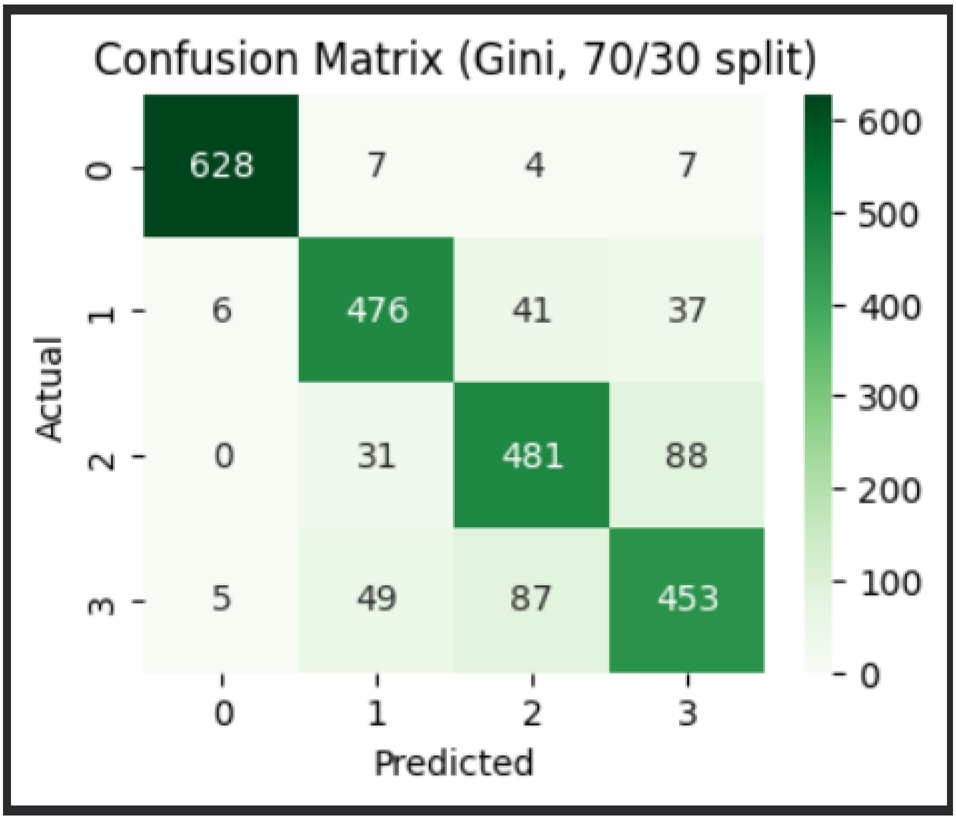
The accuracy results show that both Gini and Entropy perform well on the EEG dataset. Entropy achieved the highest accuracy (0.88125) using the 80/20 split, while Gini performed best on the 90/10 split. Overall, Entropy slightly outperformed Gini across the different data partitions.

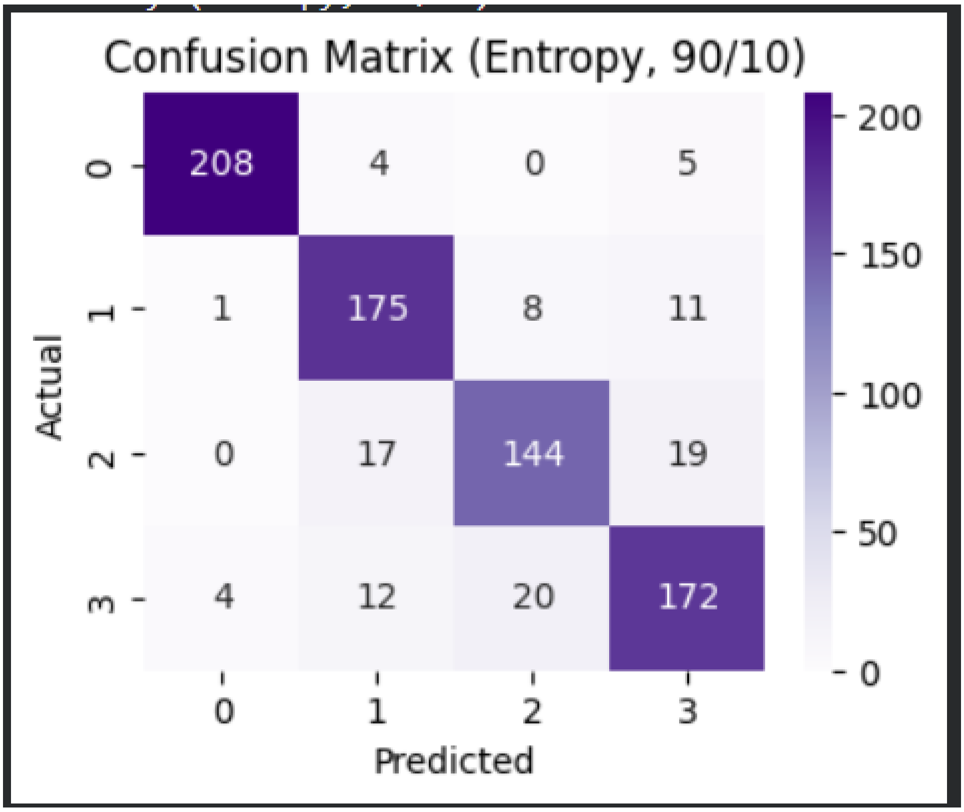
6.1.2 Confusion Matrices:

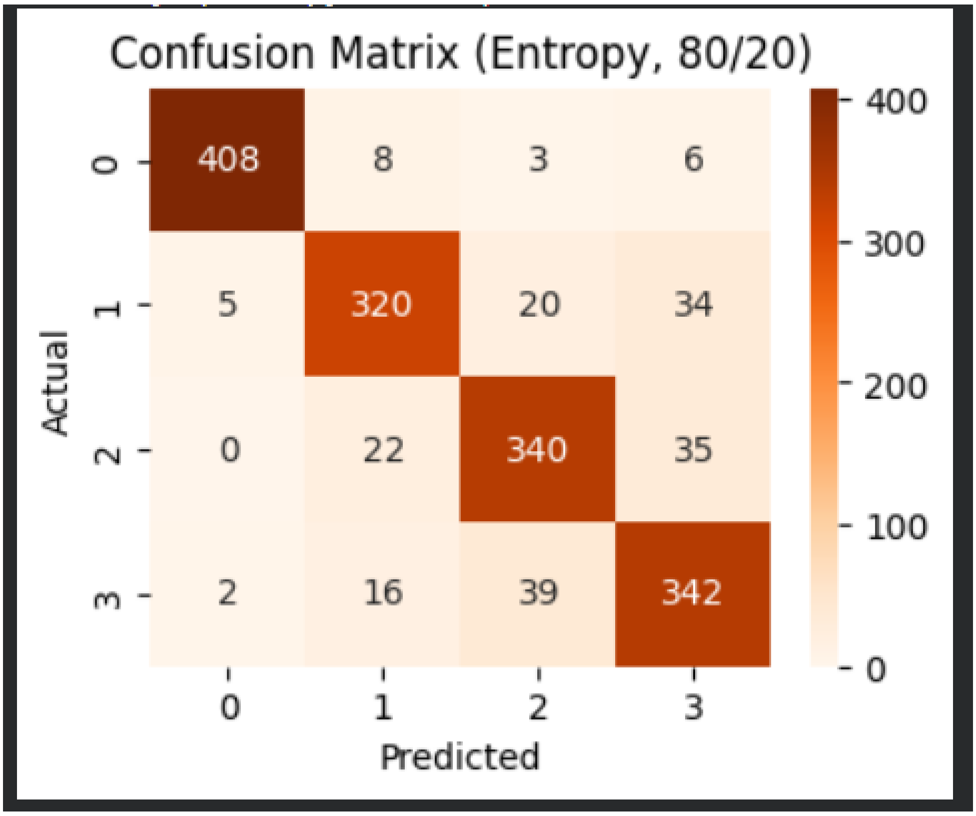
Each confusion matrix shows how well the model classified each EEG class.

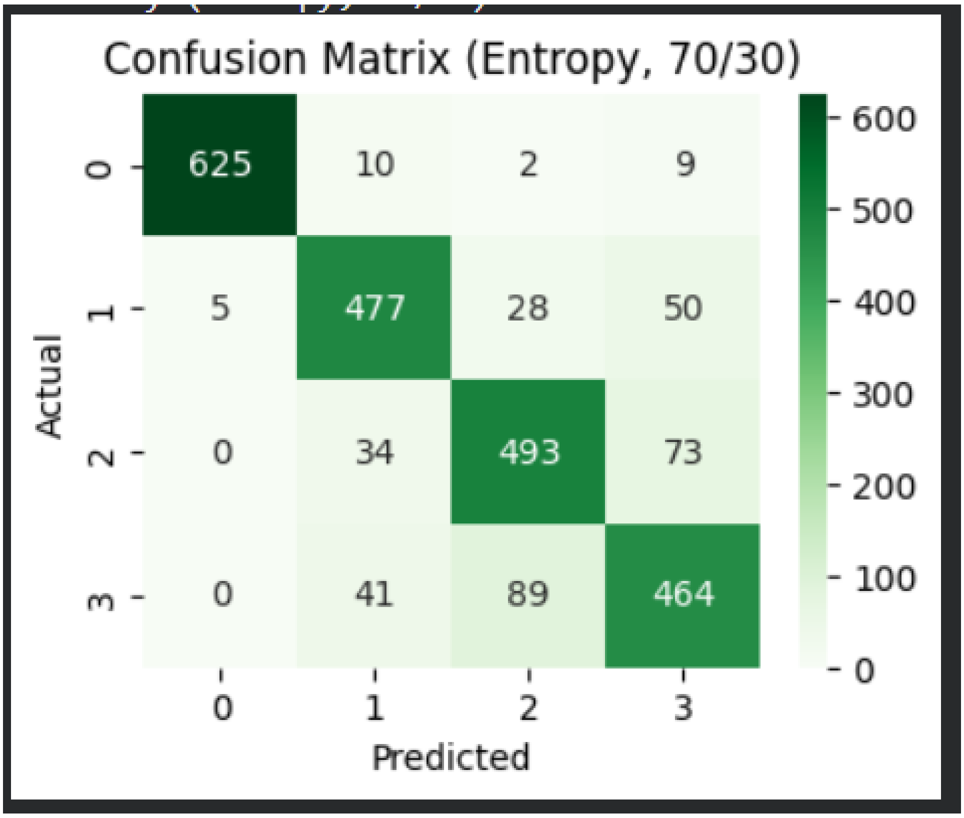

Confusion Matrix (Gini, 90/10 split)
Actual
Predicted
208
7
0
2
3
165
13
14
0
11
147
22
4
7
17
180
0
1
2
3
200
150
100
50
0


Confusion Matrix (Gini, 80/20 split)
Actual
Predicted
411
6
1
7
3
319
28
29
1
25
327
44
4
23
46
326
0
1
2
3
400
300
200
100


Confusion Matrix (Gini, 70/30 split)
Actual
Predicted
628
7
4
7
6
476
41
37
0
31
481
88
5
49
87
453


Confusion Matrix (Entropy, 90/10)
Actual
Predicted
208
4
0
5
1
175
8
11
0
17
144
19
4
12
20
172


Confusion Matrix (Entropy, 80/20)
408
8
3
6
5
320
20
34
0
22
340
35
2
16
39
342
Actual
Predicted
0
1
2
3
400
300
200
100
0


Confusion Matrix (Entropy, 70/30)
625
10
2
9
5
477
28
50
0
34
493
73
0
41
89
464
Actual
Predicted
0
1
2
3
600
500
400
300
200
100
0

6.1.3 Interpretation of Results:

Both Gini and Entropy Decision Tree models performed strongly on the EEG dataset, with accuracies above 0.84 for all splits. Entropy showed slightly better performance overall, achieving its highest accuracy (0.88125) on the 80/20 split, while Gini performed best on the 90/10 split (0.87500). This indicates that Information Gain was slightly more effective in selecting useful EEG features.

The confusion matrices show that most predictions lie on the diagonal, meaning the models classified many samples correctly—especially for Class 0 and Class 3. Most errors occurred between Class 1 and Class 2, likely due to overlapping EEG patterns between these classes.

Across the train/test splits, the 70/30 split produced the lowest accuracy for both criteria due to reduced training data, while 80/20 was the best for entropy and 90/10 for Gini. Overall, both methods captured the main EEG patterns well, with entropy showing a small advantage in generalization.

6.1.4 Feature Importance:

The Decision Tree models identified a few EEG channels as the most important for classification. In both the Gini and Entropy trees, Channel8, Channel11, and Channel7 appeared at the top levels of the tree, meaning they provided the highest reduction in impurity and entropy. These channels contributed most to separating the EEG classes. Other channels only appeared deeper in the tree, indicating lower importance. Overall, the results show that the classification mainly depends on a small set of highly informative EEG channels.