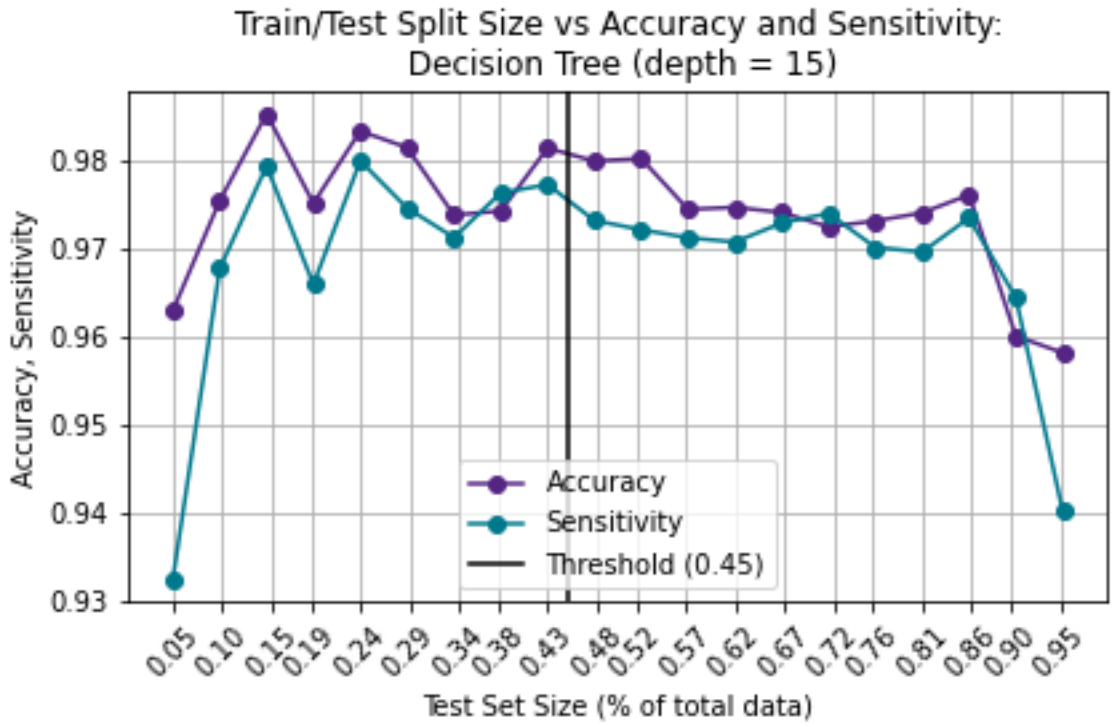
Detecting Brain Tumours using Machine Learning

The Effect of the Size of Training and Test Sets

- Threshold value for size: 0.45

Model	Accuracy	Accuracy (Size < 0.45)	Sensitivity	Sensitivity (Size < 0.45)
Logistic Regression	Accuracy: 98.1685 Sensitivity: 96.7078 Size: 0.145	Accuracy: 98.1685 Sensitivity: 96.7078 Size: 0.145	Accuracy: 98.1685 Sensitivity: 96.7078 Size: 0.145	Accuracy: 98.1685 Sensitivity: 96.7078 Size: 0.145
SVM – RBF	Accuracy: 97.2981 Sensitivity: 94.4664 Size: 0.905	Accuracy: 97.2752 Sensitivity: 94.1935 Size: 0.0975	Accuracy: 97.2527 Sensitivity: 94.6502 Size: 0.145	Accuracy: 97.2527 Sensitivity: 94.6502 Size: 0.145
SVM – Sigmoid	Accuracy: 97.1819 Sensitivity: 94.8222 Size: 0.9525	Accuracy: 96.9865 Sensitivity: 93.7164 Size: 0.335	Accuracy: 97.1819 Sensitivity: 94.8222 Size: 0.9525	Accuracy: 96.8864 Sensitivity: 93.8272 Size: 0.145
SVM – Linear	Accuracy: 97.8202 Sensitivity: 95.4839 Size: 0.0975	Accuracy: 97.8202 Sensitivity: 95.4839 Size: 0.0975	Accuracy: 97.8022 Sensitivity: 95.8848 Size: 0.145	Accuracy: 97.8022 Sensitivity: 95.8848 Size: 0.145
Decision Tree (depth = 6)	Accuracy: 98.4549 Sensitivity: 97.1631 Size: 0.43	Accuracy: 98.4549 Sensitivity: 97.1631 Size: 0.43	Accuracy: 98.0471 Sensitivity: 97.4198 Size: 0.8575	Accuracy: 98.4496 Sensitivity: 97.2500 Size: 0.24
Decision Tree (depth = 15)	Accuracy: 98.5348 Sensitivity: 97.9424 Size: 0.145	Accuracy: 98.5348 Sensitivity: 97.9424 Size: 0.145	**Accuracy: 98.3389 Sensitivity: 98.0000 Size: 0.24**	**Accuracy: 98.3389 Sensitivity: 98.0000 Size: 0.24**
Naïve Bayes	Accuracy: 97.0027 Sensitivity: 94.1935 Size: 0.0975	Accuracy: 97.0027 Sensitivity: 94.1935 Size: 0.0975	Accuracy: 96.7033 Sensitivity: 94.6502 Size: 0.145	Accuracy: 96.7033 Sensitivity: 94.6502 Size: 0.145

Best: Decision Tree (depth = 15) with test set size = 0.24

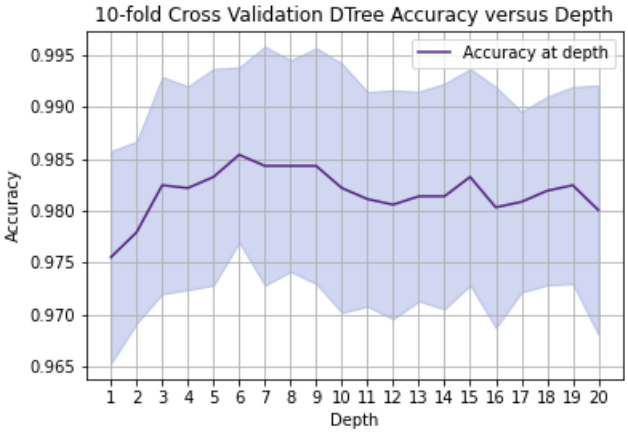


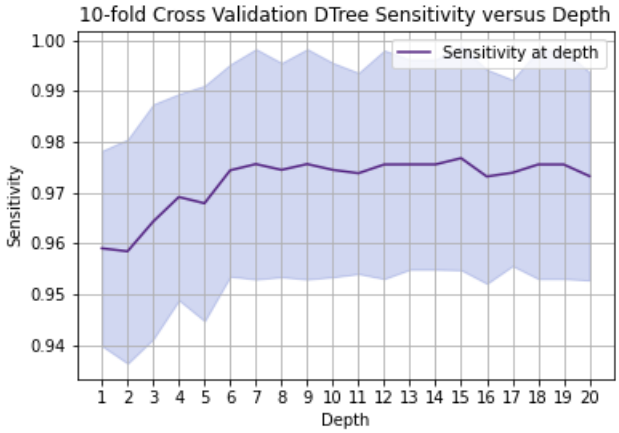
Cross-validation

- Without repetition
- With 3 repetitions

Model	Without Repetition	With 3 Repetitions
Logistic Regression	Accuracy: 97.7674 Sensitivity: 95.5951	Accuracy: 97.7674 Sensitivity: 95.6138
SVM – RBF	Accuracy: 97.3687 Sensitivity: 94.6420	Accuracy: 97.3241 Sensitivity: 94.5617
SVM – Sigmoid	Accuracy: 96.9966 Sensitivity: 93.7507	Accuracy: 97.0051 Sensitivity: 93.7516
SVM – Linear	Accuracy: 97.7940 Sensitivity: 95.7060	Accuracy: 97.8115 Sensitivity: 95.7488
Decision Tree (depth = 6)	Accuracy: 98.5380 Sensitivity: 97.4401	-
Decision Tree (depth = 15)	**Accuracy: 98.3257** **Sensitivity: 97.6792**	-
Naïve Bayes	Accuracy: 96.6776 Sensitivity: 93.8853	Accuracy: 96.6685 Sensitivity: 93.8933

Best: Decision Tree (depth = 15)





Feature Selection

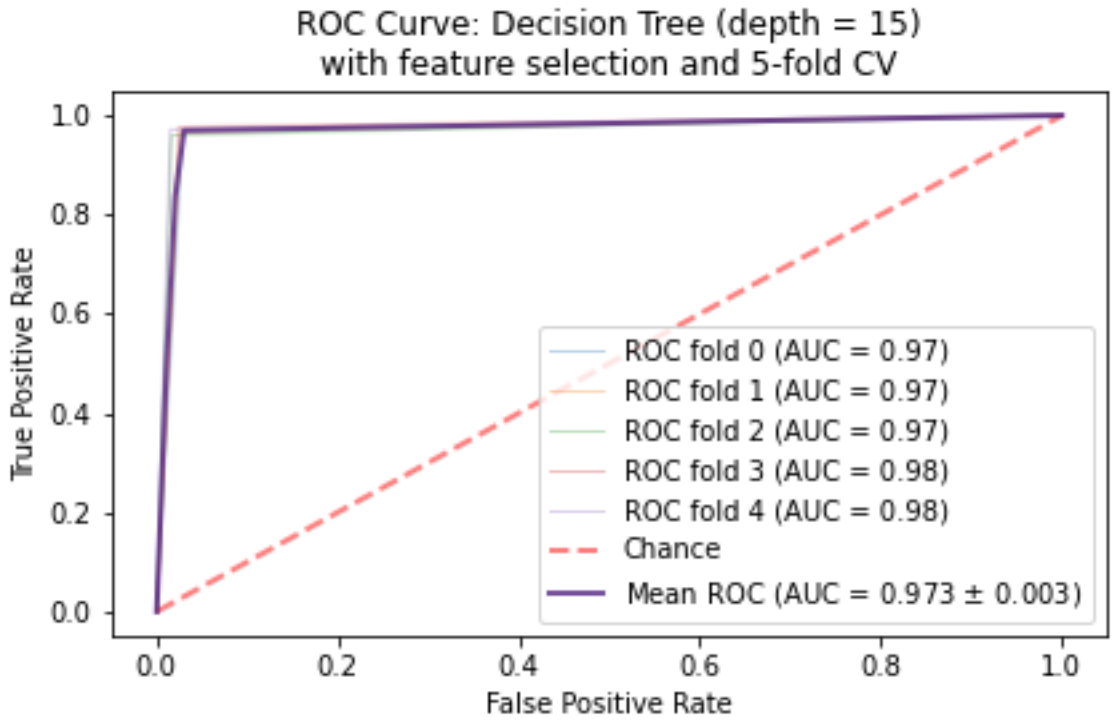
- Select 6 best using ANOVA F-value

Feature	Score
Energy	10913.5182
Homogeneity	9587.7585
Entropy	5772.6099
ASM	5086.0272
Dissimilarity	1685.2546
Skewness	727.5285

ROC Curve with 6-best feature selection and 5-fold cross-validation

Model	Accuracy	Sensitivity	Mean ROC AUC
Logistic Regression	97.5280	95.2835	0.989 ± 0.002
SVM – RBF	97.3154	94.8078	0.989 ± 0.002
SVM – Sigmoid	97.1293	94.3930	0.988 ± 0.002
SVM – Linear	97.5014	95.2202	0.989 ± 0.002
Decision Tree (depth = 6)	97.9798	96.7122	0.982 ± 0.004
Decision Tree (depth = 15)	**97.3952**	**96.8982**	**0.973 ± 0.003**
Naïve Bayes	97.3154	94.8027	0.988 ± 0.002

Best: Decision Tree (depth = 15)



Adaptive Boosting

1. Cross-validation without repetition:
   1.1. Accuracy: 98.6179
   1.2. Sensitivity: 97.6772
2. Cross-validation with 3 repetitions:
   2.1. Accuracy: 98.7506
   2.2. Sensitivity: 97.9049
3. ROC Curve with 6-best feature selection and 5-fold cross validation:
   3.1. Accuracy: 97.8998
   3.2. Sensitivity: 96.8426
   3.3. Mean ROC AUC: 0.990 ± 0.001
4. Changing number of estimators:
   4.1. Overall highest accuracy: 98.8480
      4.1.1. Corresponding sensitivity: 98.0198

4.1.2. Corresponding number of estimators: 100

4.2. Overall highest sensitivity: 98.1824

4.2.1. Corresponding accuracy: 98.8480

4.2.2.Corresponding number of estimators: 350

4.3. Selected model (similar performance but shorter training time):

4.3.1. Accuracy: 98.8303

4.3.2. Sensitivity: 98.1818

4.3.3. Estimators: 200

5. Changing ensemble learner and number of estimators:

5.1. Estimators: 50 (default for scikit-learn 0.23.2)

5.1.1. Overall highest sensitivity: 98.2261

5.1.2. Corresponding accuracy: 99.0430

5.1.3. Decision tree depth: 6

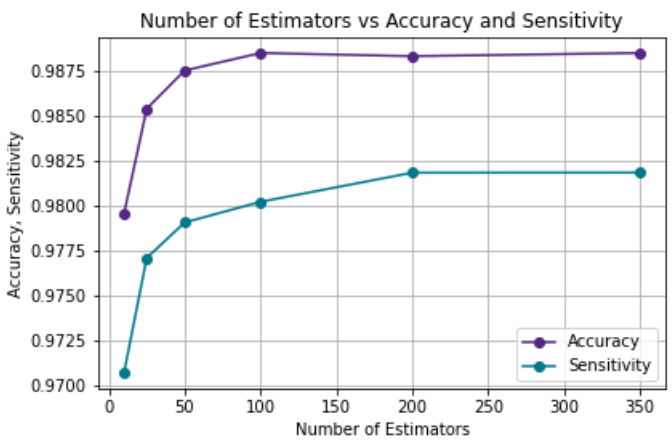
5.2. Estimators: 200

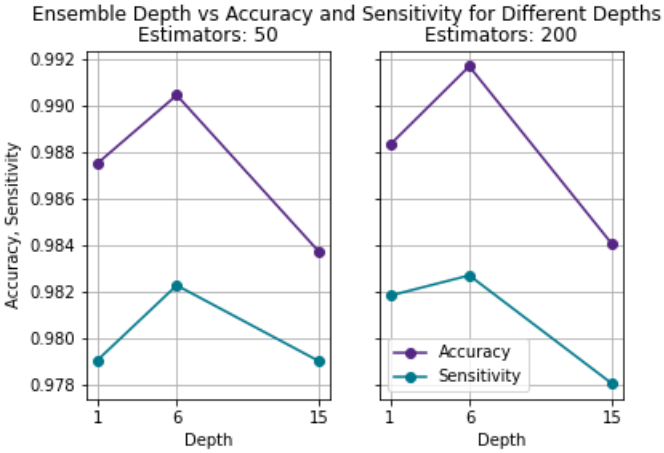
5.2.1. Overall highest sensitivity: 98.2689

5.2.2. Corresponding accuracy: 99.1671

5.2.3. Decision tree depth: 6

Best: AdaBoost with 200 estimators that use decision trees of depth 6





Convolutional Neural Net

- 3 convolutional layers
- 3 pooling layers
- 2 dropout layers
- 50 epochs
- **Accuracy: 94.95**
- **Sensitivity: 96.6**

Best overall models are the AdaBoost model with 200 estimators of trees of depth 6, and the CNN.