UNIVERSITY OF COPENHAGEN

A thesis presented to the Faculty of Science in partial fulfillment of the requirements for the degree

Bachelor Project in Machine Learning and Data Science

Automatic quality assessment tool

A tool for structural brain MRIs

Rune Brix Thomsen & Mehmet Ilker Ünsal

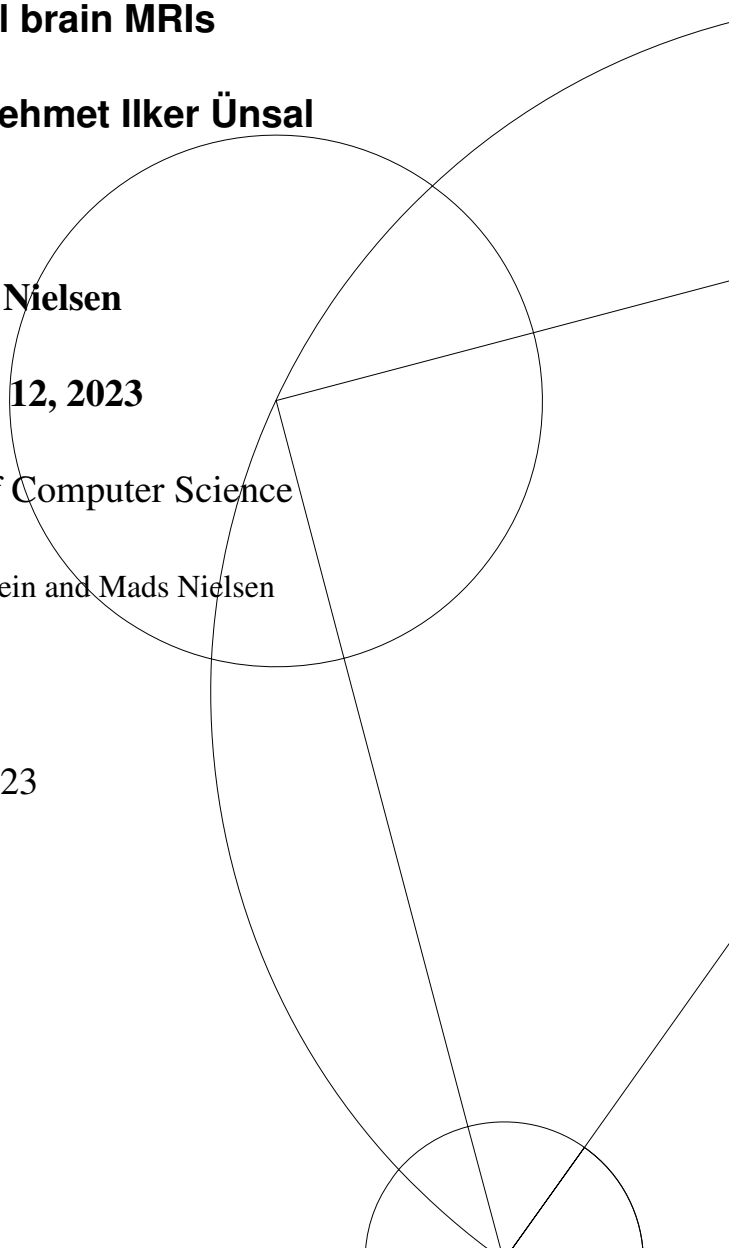
Advisor: Mads Nielsen

Submitted: June 12, 2023

Department: Department of Computer Science

Supervisor: Kiril Vadimovic Klein and Mads Nielsen

July 31, 2023



ABSTRACT

Segmentation of structural brain MRI scans plays a crucial role in medical studies and the diagnosis of neurological disorders. While recent projects have employed deep learning techniques for automated segmentation, practical implementation faces challenges. Low scan quality and significant brain pathology can hinder the effectiveness of automatic segmentation, limiting its applicability in medical studies. To address this, current research focuses on developing automatic brain segmentation validation. This paper enhances existing methods by introducing a filtering mechanism that excludes scans with scan parameters associated with poor segmentation performance, aiming to improve the quality of subsequent analyses. Our study utilizes a dataset of 45,284 3D T1 weighted MRI scans of 14893 patients harvested in the hospitals in the capital region of Denmark, accompanied by corresponding scanning parameters and annotated with the existence of major pathology in the brain and with a label for the correctness of the automated brain segmentation. Multiple machine learning models are employed to predict segmentation failure based on MRI scanning parameters. Our model successfully predicted MRI brain segmentation failure with about 85% accuracy, based only on scan parameters. This discovery enables further preprocessing of data for quality assessment, enhancing the accuracy of brain segmentation validation. Consequently, this research contributes to improved medical studies and further streamlines the brain MRI process.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

This study relies heavily on magnetic resonance imaging (MRI) scans of the human brain. MRI scans are widely utilized in medical research and the diagnosis of neurological disorders because they provide detailed visualization of internal body structures, including soft tissues and the brain. The high contrast resolution of MRI scans enables the differentiation of brain segments and the identification of abnormalities. An important advantage of MRI is its non-invasive nature, as it employs strong magnetic fields to generate images without exposing patients to radiation, unlike X-rays or CT scans. Furthermore, MRI offers the capability of acquiring multi-planar images, enhancing the accuracy of diagnoses by providing comprehensive views of anatomical structures from various angles. The significance of MRI in medical research, particularly in the early detection of neurodegenerative diseases, has been widely recognized [1]. However, one challenge associated with MRI scans is the time-consuming process of manually segmenting brain structures. To alleviate this burden on medical experts, automatic brain segmentation techniques have gained increasing attention, as they can significantly reduce the time required for analysis.

While automatic brain segmentation holds potential as a valuable tool for assisting medical professionals, there are still challenges in applying these models to scanning data. Addressing these challenges and improving the accuracy and reliability of automatic segmentation methods is crucial for their widespread adoption and effective integration into clinical practice.

1.1 PROBLEMS OF AUTOMATION

In the development of automatic brain segmentation of MRI scans, such as the work conducted by Ghazi et al. [2], it has been observed that certain scans fail to be segmented accurately. Consequently, ensuring quality control of automated segmentation becomes imperative before the application of these techniques in medical practice. In this context, another valuable application of automation arises from the automatic quality assessment of MRI scans. Automatic assessments of the quality of brain-segmented scans are essential to further enhance the effectiveness and practicality of automation. In previous investigations of failed segmentations, significant brain pathologies were identified as influential factors contributing to segmentation failures [1, p.2-3]. However, even in the absence of major pathologies, some scans still exhibited inadequate segmentation results. Consequently, exploring additional factors can provide valuable insights into the circumstances under which segmentation models fail to accurately segment brain MRIs.

1.2 USAGE OF SCAN PARAMETERS

Numerous studies have been conducted on the automation of quality control in MRI scans. For instance, Bottani et al[3] employed data annotation techniques to identify inadequate T1-weighted brain MRIs, detect gadolinium injection, and evaluate image quality. Although the first two objectives were accomplished with precision, the overall image quality rating exhibited some inaccuracies. Similarly, Alfaro-Almagro et al[4] manually classified datasets by categorizing problems and generating features for quality control. These features were subsequently inputted into a supervised learning classifier. While these studies have achieved some success in automating quality control, their approaches necessitate significant preprocessing and fail to explore the practical applicability of this quality

control in assessing the accuracy of subsequent automated brain segmentation. To bridge this gap, our study aims to evaluate the image quality of MRI scans based solely on the metadata associated with the scans. The quality of MRIs is influenced by various factors, including patient-specific and scanning-specific parameters. Parameters such as patient positioning during scanning, the type of scanning machine employed, and image resolution all impact the quality of MRI scans. It remains unclear whether certain scans can be selectively excluded from automated segmentation based solely on scan parameters. Therefore, our research endeavours to investigate the feasibility of selectively omitting scans from the dataset intended for automatic segmentation, thereby enhancing the reliability of the automated segmentation process.

1.3 MACHINE LEARNING AND PATTERN FINDING

Machine learning models offer valuable insights into patterns and specific parameters that impact the image quality of MRI scans. This is because machine learning uses mathematical concepts and algorithms to analyse large datasets and identify complex relationships that are not necessarily apparent through more traditional analytical approaches. One significant advantage of machine learning is its ability to capture non-linear relationships between features, such as scan parameters, and the corresponding output, such as segmentation failure. Machine learning methods can effectively model non-linear relationships, enabling the identification of intricate patterns and interactions within the data.

1.4 DATA PREPARATION

Data preparation is a crucial step in the development of any machine learning classifier, and this study recognized its importance in ensuring reliable and consistent data for training and testing purposes. Several reasons highlight the significance of data preparation:

- 1. Ensuring Data Quality: Data preparation involves addressing various issues related to data quality, such as handling missing values, removing duplicate features, and dealing with outliers. These steps help to enhance the reliability and consistency of the data used in the analysis.
- 2. Transforming Data for Machine Learning Algorithms: Raw data often requires transformation to make it suitable for machine learning algorithms. This process, known as feature engineering, may involve scaling, normalization, encoding, or other transformations that create more informative data for the machine learning algorithm to learn from.
- 3. Integrating and Standardizing Data Sources: Data may originate from diverse sources with varying structures and quality. Data preparation involves integrating different data sources to create a unified dataset that is compatible and consistent across all aspects of its usage. This step ensures that the data is harmonized and ready for training the machine learning models.

In the specific context of this study, the inclusion of string data in the dataset presented a significant challenge in data preparation. Machine learning algorithms operate on numerical data, and as strings lack inherent numerical representations, they cannot be directly used for calculations. Therefore, all string data needed to be encoded into numerical representations.

By transforming the string data into numerical features, it becomes possible to leverage machine learning algorithms and utilize them for predictive modelling and analysis.

2

METHODS

The section of this thesis concentrates on the data utilized in the study, encompassing data preparation and subsequent model development. The primary focus is to investigate the feasibility of predicting segmentation failure based on scan parameters. The prediction of segmentation failure based on brain volumes will also be described. Considerable emphasis is placed on the data preparation process for scan parameters, as it constitutes a significant aspect of the study.

2.1 DATA DESCRIPTION

Data utilized in this study consisted of 45,284 3D T1 weighted MRI scans of 14893 patients harvested in the hospitals in the capital region of Denmark. As children's brains evolve quickly, scans of the brain of children were excluded from the dataset, resulting in a dataset comprising 33,426 MRI scans of 11954 patients. A more concise description of the data can be found in Klein[5, p. 16-19]
All brain MRIs were segmented in 132 distinct brain regions using FAST-AID Brain segmentation[2]. A subset of randomly selected brain MRIs was annotated by a neurologist based on raw MRI scans and the corresponding segmentations with two labels each. The labels indicate the presence and severity of brain pathology (none, intermediate, severe) and the degree of segmentation failure (none, intermediate, severe).

2.1.1 *Brain volumes*

For the dataset to be used for predicting segmentation failure based on brain volumes, each scan was characterized by 137 features. 132 volumes for the aforementioned brain regions and five features consisting of two identification-related features, a column indicating background and the two aforementioned labels.

To ensure the suitability of volumes for training a machine learning model, certain scans had to be excluded from consideration. Scans that presented an ”intermediate” value in the label columns were excluded along with the column for the background feature. To prevent bias in our analysis, only a single scan from each patient was included in the model training process.

Consequently, the resulting dataset comprised 2744 brain volumes, encompassing 132 features after excluding the identification-related features. Among these volumes, 1738 were categorized as segmentation failures, while 1006 were classified as correctly segmented. The dataset was subsequently divided into a test set, comprising 533 volumes, and a training set, comprising 2211 volumes. A contingency table for the dataset can be seen in table 2

Brain Segmentation\Major pathology	0	1	Total
0	1003	3	1006
1	387	1351	1738
Total	1390	1354	2744

Table 2: Contingency table for dataset

2.1.2 *Scan parameters*

A dataset of the scanning parameters for the same scans as the one described in the brain volumes was used to determine segmentation failure based on scanning parameters. The dataset included 2744 MRI scans with 50 scanning parameters and 2 ID features. These parameters adhere to the Digital Imaging and Communications in Medicine (DICOM) standard format, which serves as a universal standard for representing medical imaging data. For a comprehensive overview of DICOM features, please consult the Innolitics DICOM metadata reference [6]. From these 52 features, a subset of parameters was selected to be eligible for use in predictions. The features excluded were determined to provide no additional insight into the circumstances in which the scans had taken place. A list of the included scanning parameters can be found in figure 9. The dataset also made use of the binary label values that serve as indicators for classifying whether a scan had been correctly or incorrectly segmented, as well as whether the scanned brain exhibited a major pathology.

The resulting dataset contained 24 scanning parameters, 2 ID features and 2 binary labels.

2.2 ENCODING METHODS AND DATA CLEANING OF SCANNING PARAMETERS

In order to prepare the scanning parameter data for the development of machine learning models for brain segmentation classification, several preprocessing steps were undertaken. Firstly missing values were identified in certain scan parameters within some scans. To address this issue, the missing values were filled using either the mean or median values. The choice between the mean and median was made by considering the nature of the values within the respective column. If the values exhibited minimal deviation from fixed values, the median was used to avoid introducing entirely new values. Conversely, if the values showed deviation from each other, the mean was employed to capture a better spread of the values.

Thirdly, the metadata in the DICOM data encompassed both numerical and string values. However, the machine learning models utilized in testing were unable to process string values during model development. Consequently, various encoding methods were employed to transform the dataset before training the models. The primary objective of encoding was to accurately represent string values as numerical values within the dataset. The first method involved identifying string columns that exhibited multi-categorical-like behaviour. These columns consisted of arrays of strings that described specific features included in a particular parameter. To represent these strings accurately as numerical values indicating the presence of a particular feature in a scan, a multi-label binarizer encoding technique was applied. This encoding transformed the multi-categorical string values in the columns into binary matrices, intuitively indicating the presence or absence of a label [7]. The second method involved handling columns that contained a single categorical string. In these cases, a label encoder was employed to assign a unique numerical value to each category. This encoding process replaced the string values in each column with their corresponding numerical representations[8].

The final step in the data preparation phase involved removing columns that exhibited the same value in 99% of the scans. These columns were deemed to offer no additional value in terms of pattern recognition for the machine-learning models. The resulting dataset contained 2744 MRI scans each with 84 features including 2 ID features and 2 label features. Before the dataset was deployed in the training of a machine learning model, the indication of whether a patient has a major pathology would be utilized to filter out scans where the patient had a major brain pathology, as these scans were more likely to exhibit segmentation failures. After excluding scans where the patient had a major brain pathology, the class balance ratio was calculated and applied to the data by dividing the number of labels minus the sum of the labels by the sum of the labels.

2.3 MACHINE LEARNING MODELS FOR CLASSIFICATION

A central part of the study was to utilize machine learning models to assert the correctness of the automated segmentation by looking at brain volumes and to improve this ability by trying to find potential patterns in scan parameters that would indicate that automatic segmentation had failed. A wide variety of models were included in the testing. The models used can be seen in Table 3. The

Name	Description
XGBoost	Employs gradient boosting on an ensemble of weak decision tree models. Each tree makes predictions based on the weighted features and minimises a loss function through the use of gradient descent
Random Forest	Constructs multiple decision trees and outputs predictions based on the majority vote. Each tree is typically a binary tree, where each internal node splits the data based on a specific feature and threshold.
Neural Network	Uses nodes in layers to learn complex patterns by adjusting connection strength between nodes. The layers of a Neural network consist of an input layer, one or more hidden layers responsible for learning patterns, and an output layer
KNN	Uses a distance metric to measure the similarity between instances in the feature space and uses the majority vote to determine the class label for new instances.

Table 3: Machine learning models deployed for testing

primary objective of this study was to evaluate the performance of several machine learning models in identifying patterns within scan parameters. The XGBoost model was implemented using the xgb function from the xgboost library [9], with parameters set to 100 trees, a maximum depth of 3, and a learning rate of 0.1. The Random Forest model was implemented using the Random Forest classifier from the sklearn library [10], with an ensemble of 100 trees.

Two different neural network models were used to determine the failure of brain segmentation based on brain volumes and scan parameters. The brain volume model consisted of four layers: an input

layer, a dropout layer with a rate of 0.5, a hidden layer with 16 units, and an output layer with 1 unit. The input and hidden layers used “relu” activation functions, while the output layer used a “sigmoid” function.

The scan parameter model consisted of three layers: an input layer, a dropout layer with a rate of 0.5, and an output layer with 1 unit. The input layer used a “relu” activation function, while the output layer used a “sigmoid” function.

Both models were trained using the Adam optimizer and binary cross-entropy loss for 100 epochs with a batch size of 32. The scan parameter model included a validation split of 0.2, while the brain volume model did not [11].

The sklearn library KNeighborsClassifier was used to implement the KNN model. The parameters used were the default parameters[12].

2.4 VALIDATION AND EVALUATION METHODS

A comprehensive range of accuracy measurement techniques was employed to gauge the models’ success. These techniques encompassed the creation of a confusion matrix, calculation of sensitivity and specificity, and calculation of Precision-Recall AUC and ROC-AUC.

The confusion matrix specifies a table where each row represents a class and where each column represents the instances in a predicted class. An example of a confusion matrix can be seen in table 4: Accuracy of the models developed in this study are calculated using the performance metric seen

	Predicted Positive (PP)	Predicted negative (PN)
Positive (P)	True positive (TP)	False negative (FN)
Negative (N)	False positive (FP)	True negative (TN)

Table 4: Confusion matrix

in table 5 The ROC curve is then calculated by plotting the sensitivity against the fall-out and the

Metric	Calculation
Accuracy	$\frac{TP+TN}{P+N}$
Sensitivity (Recall/True Positive Rate)	$\frac{TP}{P}$
Specificity	$\frac{TN}{N}$
Precision (Positive Predictive Value)	$\frac{TP}{TP+FP}$
Fall-Out (False Positive Rate)	$\frac{FP}{P}$

Table 5: Model Performance Metrics

Precision recall curve is then calculated by plotting the precision against the recall. Taking the area under the curve of these two curves gives us the PR-AUC and ROC-AUC[13].

In addition to accuracy measurement, particular attention was dedicated to exploring the significance of individual features within the models. Various approaches were employed to assess feature importance. For the neural network, the feature importance was calculated by analyzing the weights of the connections between the input layer and the first hidden layer. We then took the absolute values of these weights and computed the mean along each row to obtain the feature importances. The Random Forest model's feature importance scores are calculated based on the information gain or decrease in impurity each feature provides when it is used for splitting in the decision trees within the random forest. XGBoost uses the same method when calculating feature importance. The KNN classifier estimates feature importance by examining the relevance of features based on how frequently they are chosen among the nearest neighbours. We then normalize the importance scores by dividing each score by the sum of all scores to obtain relative importance.

3

RESULTS

3.1 BRAIN VOLUMES CLASSIFICATION

The results of the classification of failure of brain segmentation based on brain volumes demonstrated that XGBoost, Random Forest, and Neural Network models consistently achieved high accuracy, ROC-AUC, PR-AUC, sensitivity, and specificity scores, all exceeding 0.96. An overview of the models' performance can be found in table 7.

In particular, the Neural Network and Random Forest models emerged as the best-performing models, with near-perfect scores across all metrics. Both models achieved accuracy, ROC-AUC, PR-AUC, sensitivity, and specificity values of 0.97 and above.

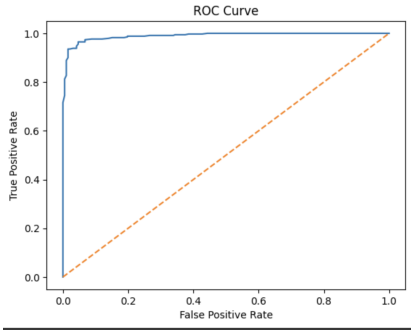
Conversely, the KNN model exhibited comparatively lower performance. While it achieved an accuracy of 0.83, the KNN model displayed lower values in other metrics such as ROC-AUC, PR-AUC, sensitivity, and specificity.

The graphs in figure 1 - 4 show the ROC and PR curve for each respective classification model of brain segmentation assessment based on brain volumes. Each ROC curve shows an orange dotted line which represents the random guess for each model while the blue line is the actual performance of said model. For the PR models, the area under the curve has been included at the top of each model, this number is equal to the PR-AUC value found in table 7.

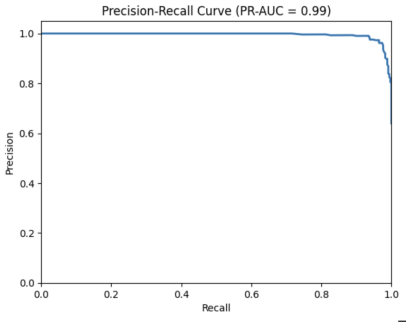
Table 6: Brain volumes model performance metrics

Model	**Accuracy**	**ROC-AUC**	**PR-AUC**	**Sensitivity**	**Specificity**
XGBoost	0.96	0.99	0.99	0.96	0.94
Random Forest	0.97	0.99	0.99	0.97	0.97
Neural Network	0.97	0.99	0.99	0.97	0.97
KNN	0.83	0.93	0.97	0.77	0.95

Table 7: Brain volumes model Performance Metrics

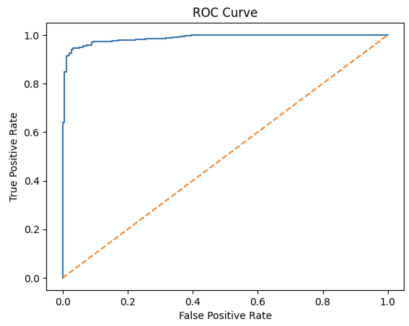


(a) Random forest ROC for brain volumes

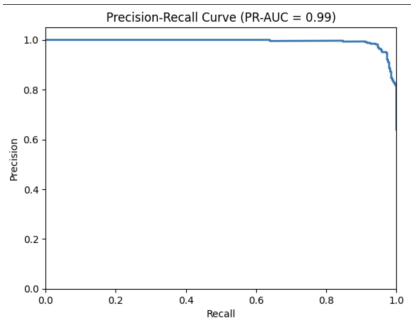


(b) Random forest PR for brain volumes

Figure 1: Random forest graphs for brain volumes.

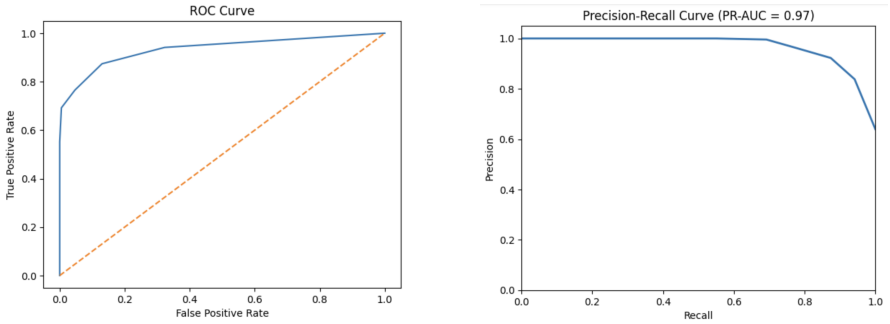


(a) XGBoost ROC for brain volumes



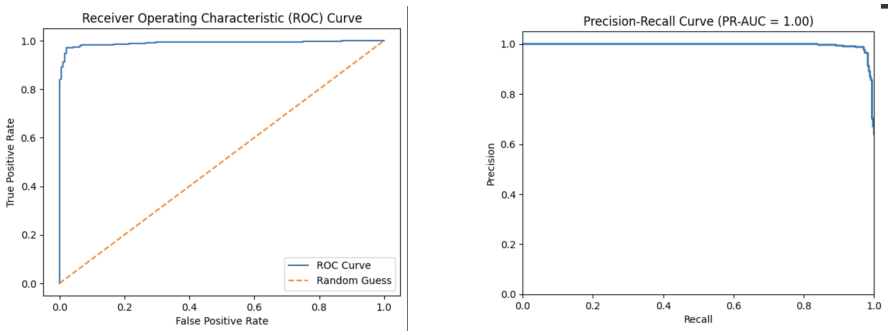
(b) XGBoost PR for brain volumes

Figure 2: XGBoost graphs for brain volumes



(a) KNN ROC for brain volumes (b) KNN PR for brain volumes

Figure 3: KNN graphs



(a) NN ROC for brain volumes (b) NN PR for brain volumes

Figure 4: NN graphs for brain volumes

3.2 SCANNING PARAMETERS CLASSIFICATION

The results of the classification of failure of brain segmentation based on scan parameters demonstrate that all models achieved accuracy values ranging from 0.84 to 0.86. In addition, the models' performance across ROC-AUC and PR-AUC scores ranged from 0.82 to 0.91. An overview of the models' performance can be found in table 8.

Among the models evaluated, the Neural Network exhibited the most consistent performance. It achieved an accuracy of 0.85, along with ROC-AUC and PR-AUC scores of 0.90 and 0.86, respectively. Furthermore, the Neural Network demonstrated sensitivity and specificity values of 0.79 and 0.87.

Of the models, the KNN (K-Nearest Neighbors) algorithm exhibited relatively weaker performance

compared to the other models. While it achieved an accuracy of 0.86, the KNN model demonstrated lower values in other metrics. It showed a lower ROC-AUC score of 0.87, indicating lower accuracy in distinguishing between classes. Additionally, the KNN model yielded a PR-AUC score of 0.82, suggesting slightly reduced precision and recall performance. The sensitivity value of 0.69 indicates a relatively lower ability to correctly identify positive instances. However, the KNN model did showcase a specificity of 0.92, indicating a stronger ability to correctly identify negative instances.

Analysis of the top 5 feature importances across different models reveals consistent patterns. The

Model	**Accuracy**	**ROC-AUC**	**PR-AUC**	**Sensitivity**	**Specificity**
XGBoost	0.85	0.91	0.91	0.78	0.87
Random Forest	0.84	0.89	0.86	0.77	0.87
Neural Network	0.85	0.90	0.86	0.79	0.87
KNN	0.86	0.87	0.82	0.69	0.92

Table 8: Model Performance Metrics

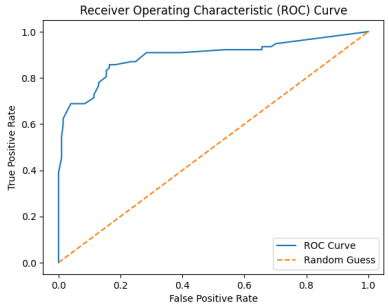
SeriesDescription attribute appears prominently, indicating its significance. Features such as 'SeriesDescription_con', 'SeriesDescription_gd', and 'SeriesDescription_k' consistently rank high in importance across XGBoost, Random Forest, Neural Network, and KNN models. This suggests a strong predictive relationship with the target variable. The presence of 'SequenceVariant_MP' and 'SequenceVariant_SP' in the top feature importances highlights the role of sequence variations. A full overview of the top five important features for each model can be found in table 9.

The graphs in figure 5 - 8 show the ROC and PR curve for each respective classification model of

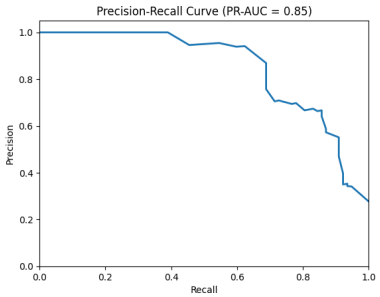
XGBoost	**Random Forest**	**Neural Network**	**KNN**
SeriesDescription_k: 0.1169	SeriesDescription_con: 0.2735	SeriesDescription_k: 0.0411	SeriesDescription_con: 0.0604
SeriesDescription_con: 0.1113	SeriesDescription_gd: 0.1308	SeriesDescription_con: 0.0392	SeriesDescription_gd: 0.0284
SequenceVariant_MP: 0.1096	SeriesDescription_k: 0.0365	SeriesDescription_gd: 0.0369	SeriesDescription_t1: 0.0223
SeriesDescription_gd: 0.1010	SequenceVariant_MP: 0.0364	SeriesDescription_c: 0.0341	SeriesDescription_sag: 0.0122
ScanOptions_OTHER: 0.0869	SequenceVariant_SP: 0.0358	SeriesDescription_siemens: 0.0241	ImageType_DIS2D: 0.0097

Table 9: Top 5 Feature Importances for Different Models

brain segmentation assessment based on scan parameters. As with the graphs in 3.1 ROC curves show an orange dotted line which represents the random guess for each model while the blue line is the actual performance of said model. For the PR models, the area under the curve has been included at the top of each model, this number is equal to the PR-AUC value found in table 8.

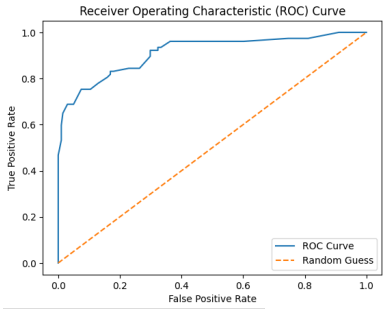


(a) Random forest ROC for scan parameters

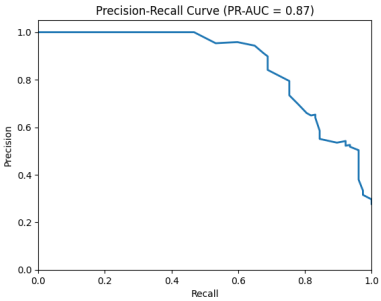


(b) Random forest PR for scan parameters

Figure 5: Random forest graphs for scan parameters

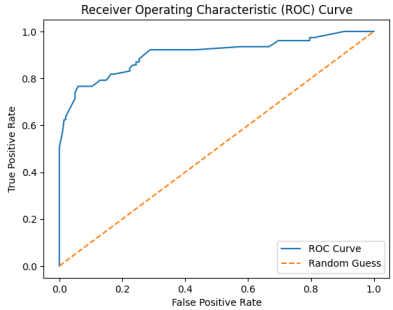


(a) XGBoost ROC for scan parameters

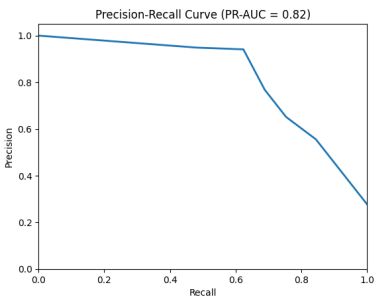


(b) XGBoost PR for scan parameters

Figure 6: XGBoost graphs for scan parameters

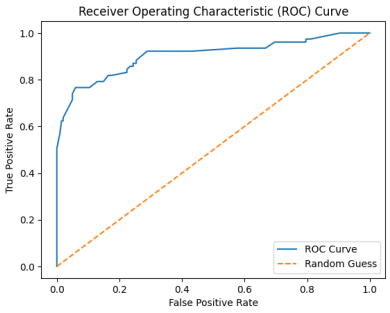


(a) KNN ROC for scan parameters

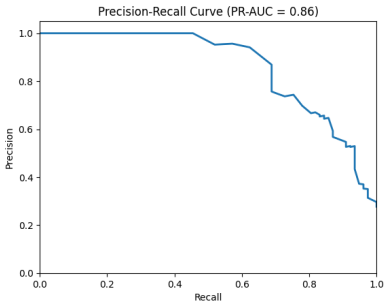


(b) KNN PR for scan parameters

Figure 7: KNN graphs for scan parameters



(a) NN ROC for scan parameters



(b) NN PR for scan parameters

Figure 8: NN graphs for scan parameters

The full code used in this study can be found at [14].

4

DISCUSSION

Brain volumes used in this study were obtained by applying the FAST-AID algorithm [2] for automatic brain segmentation on MRI scans from the Capital Region of Denmark. The corresponding scan parameters were then used to evaluate brain segmentation based on various scan parameters.

Predictive models based on brain volumes showed exceptional performance, with accuracy levels exceeding 97% in baseline models. This result confirms the feasibility of evaluating the accuracy of brain segmentation. However, these models do not provide insight into the reasons for segmentation failures, necessitating further investigation into scan parameters.

Our analysis also showed high performance (approximately 85%) when examining scan parameters alone, indicating the potential to identify cases where automatic brain segmentation using FAST-AID is likely to fail. Certain attributes within the SeriesDescription parameter, such as “con,” “k,” and “gd,” were found to be most influential in classification. However, interpreting these attributes can be challenging due to non-standard descriptors used in hospitals. Additionally, “SequenceVariant,” “ScanOptions,” and “ImageType” were identified as significant features in the models and were encoded using multi-label binarization, indicating their importance in determining the efficacy of FAST-AID in automatically segmenting brain MRI scans.

A thorough evaluation of the models also revealed that the sensitivity score of the scan parameter models ranged from 69% to 78% and a specificity of 87% to 92%, indicating that the models had

more difficulty classifying scans with brain segmentation failure than those with accurate segmentation. This suggests that scan parameters are better suited for approving brain MRI scans for automatic segmentation rather than determining if scans are unsuitable for such processing.

Data cleaning and encoding of the scan parameters were a large part of this study and therefore also a point of contention. The methods used in the study were chosen based on the impact that the different methods tested had on the final model accuracy. Methods contested included; using PCA to reduce the dimensionality of the SeriesDescription attributes, using an Autoencoder to encode all of the data before training, and using the dirty_cat similarity encoder[15] on the SeriesDescription parameters instead of a multi-label binarizer. Not only did utilising these methods not outperform the final chosen method, but the final method of using multi-label binarizer without autoencoder or PCA resulted in a more transparent dataset, allowing us to gain better insight into important parameters with the use of feature importance evaluation methods. An overview of the accuracy of the models when using the other methods can be found in figure 10-12.

5

CONCLUSION

This study demonstrated the necessary steps to perform a retrospective study on the prediction of segmentation failure in brain volumes obtained from the FAST-AID automatic brain segmentation. Despite achieving high accuracy in classifying correct and incorrect segmentations, the study was unable to provide complete accuracy or insights into the reasons behind the failures.

Therefore the study showcased the feasibility of predicting segmentation failure by considering the scan parameters of the MRI scans used with the FAST-AID algorithm. A framework was introduced to represent metadata for 3D T1-weighted brain MRIs in machine learning models, with the aim of assessing the accuracy of automatic brain segmentation. The results demonstrated a promising accuracy of approximately 85%, indicating that scan parameters contain valuable contextual information for identifying suitable 3D T1-weighted brain MRIs for automatic segmentation using FAST-AID.

Incorporating these scan parameters into the segmentation model allows for a more comprehensive and context-aware framework for predicting the suitability of 3D T1-weighted brain MRIs for automatic segmentation. This advancement has great potential for improving the overall performance and reliability of the segmentation process.

6

FUTURE WORK

After establishing a framework for representing metadata from 3D T1-weighted brain MRIs and evaluating the accuracy of automatic brain segmentation with FAST-AID, the next step is to use the findings of this study to improve the assessment of MRI quality. One approach is to employ the methodology described in this study to preprocess data intended for automatic segmentation using FAST-AID. By incorporating the preprocessing steps outlined in this research, the overall quality and suitability of brain MRIs for segmentation could be enhanced.

Future work aimed at improving the brain MRI process should further investigate the metadata elements identified as crucial in evaluating the quality of brain MRI scans. Although specific features have consistently demonstrated importance across various machine learning models, their precise implications remain challenging to interpret. An in-depth investigation of these features could provide valuable insights, leading to advancements in the automation process and further improvements in the accuracy and reliability of brain MRI segmentation.

7

APPENDIX

```
"newID", "newPatientID", "AcquisitionDuration", "dBdt" , "EchoTime", "EchoTrainLength",
"EchoNumbers", "FlipAngle", "ImagingFrequency", "MagneticFieldStrength",
"RepetitionTime", "Rows", "Columns", "SliceThickness", "SpacingBetweenSlices",
"RowSpacing", "ColumnSpacing", "DistanceBetweenSlices", "NumberOfSlices",
"SeriesDescription_x", "ImageType", "ScanningSequence", "SequenceVariant",
"ScanOptions", "Sequence", "PatientPosition"
```

Figure 9: list of parameters included in scan parameters dataset

	Accuracy	ROC-AUC	PR-AUC	Sensitivity	Specificity
XGBoost	0.74	0.78	0.78	0.60	0.79
Random Forest	0.74	0.78	0.65	0.58	0.80
Neural Network	0.71	0.76	0.65	0.70	0.72
KNN	0.79	0.76	0.66	0.44	0.93

Figure 10: Accuracy of models using autoencoder and the dirtycat encoder on the SeriesDescription parameter

	Accuracy	ROC-AUC	PR-AUC	Sensitivity	Specificity
XGBoost	0.85	0.90	0.90	0.74	0.89
Random Forest	0.86	0.89	0.85	0.74	0.90
Neural Network	0.81	0.87	0.85	0.79	0.82
KNN	0.84	0.85	0.81	0.66	0.91

Figure 11: Accuracy of models using autoencoder and PCA to reduce dimensionality on the SeriesDescription parameter after using multi-label binarizer

	Accuracy	ROC-AUC	PR-AUC	Sensitivity	Specificity
XGBoost	0.72	0.77	0.77	0.69	0.73
Random Forest	0.73	0.77	0.65	0.58	0.79
Neural Network	0.73	0.77	0.65	0.66	0.75
KNN	0.79	0.72	0.66	0.44	0.93

Figure 12: Accuracy of models without using autoencoder and using the dirtycat encoder on the SeriesDescription parameter

BIBLIOGRAPHY

1. Nagaraj Yamanakkanavar, J. Y. C. & Lee, B. MRI segmentation and classification of human brain using deep learning for diagnosis of Alzheimer's disease: a survey (2020).

2. Ghazi, M. M. & Nielsen, M. FAST-AID Brain: Fast and Accurate Segmentation Tool using Artificial Intelligence Developed for Brain. eprint: arXiv:2208.14360 (2022).

3. Bottani, S. *et al.* Automatic quality control of brain T1-weighted magnetic resonance images for a clinical data warehouse. *Medical Image Analysis* **75,** 102219. ISSN: 1361-8415. https://www.sciencedirect.com/science/article/pii/S1361841521002644 (2022).

4. Alfaro-Almagro, F. *et al.* Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage* **166,** 400–424. ISSN: 1053-8119. https://www.sciencedirect.com/science/article/pii/S1053811917308613 (2018).

5. Klein, K. V. Brain Volumetric Changes due to COVID-19. A Retrospective Study on a Danish Cohort (2022).

6. Innolitics. *DICOM standard browser* https://dicom.innolitics.com/ciods/mr-image/mr-image. (accessed: 28-05-2023).

7. learn, S. *MultiLabelBinarizer-documentation* https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MultiLabelBinarizer.html. (accessed: 28-05-2023).

8. learn, S. *LabelEncoder-documentation* https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html. (accessed: 28-05-2023).

9. XGBoost. *XGBoost-documentation* https://xgboost.readthedocs.io/en/stable/parameter.html. (accessed: 31-05-2023).

10. sklearn.ensemble.RandomForestClassifier. *sklearn.neighbors.KNeighborsClassifier* https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html. (accessed: 31-05-2023).

11. Keras. *The Sequential class* https://keras.io/api/models/sequential/. (accessed: 31-05-2023).

12. Sklearn. *sklearn.neighbors.KNeighborsClassifier* https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html. (accessed: 31-05-2023).

13. Joseph K. Blitzstein, J. H. in, 45–59 (2019).

14. Thomsen, R. & Ünsal, M. *Assessment-of-brain-segmentation-using-scan-parameters* https://github.com/RuneBrix/Assessment-of-brain-segmentation-using-scan-parameters. (accessed: 12-06-2023).

15. Dirty$_c$*at* developers. *dirty$_c$at : machinelearningwithdirtycategories* https://github.com/dirty-cat/dirty_cat. (accessed: 09-05-2023).