# Data Mining in Brain Tumor Detection

ST. JOHN'S UNIVERSITY

Tariere Timitimi, Ana Sytina, Tahmidul Haque

## Abstract

- Brain Tumors became greatly complicated due to later detections resulted in limited time for response
- Many fields rely on Data Mining Techniques to increase precision, diagnosis due to technical advancements
- Data Mining Techniques can increase early detections and speed up treatments and increase prevention rates
- Our Goal is to determine what is causing tumors to form within a human
- We can conduct hypothesis for timing, detections, and treatment plannings
- Enforcing techniques such as Predictive Modeling, Data Wrangling, and Model Evaluations

## Introduction

- Brain tumors, like other cancers can spread throughout the human body, not just the brain and/or spinal cords
- Formation of Brain Tumors are always Unpredictable and Time worthy
- Time determines how we analysis diagnosis and therefore seek treatment plans needed
- Analytical experts sometimes require so much information, some of which are difficult to obtain or uncover
- Early detections allow us to seek treatment plans and reduce the risk of the tumor faster and more efficiently
- Were enforcing MRI images through applying various data mining techniques such as Decision Trees, SVMS, Neural Networks, CNN's, etc
- Were hopeful to accurately detect possible tumors using these various Data Mining techniques

## Literature Review

- Deep Learning Models have found to provide accuracy rate of 99.5% in detecting various Tumors
- Models such as CBAM, SPPF+, and BiFPN all play crucially in accurately determine what type of tumor is specifically present
- Image Preprocessing also played a role as it eliminates and noisy or unwanted elements within the images.
- Various Data Mining techniques and Models were enforced which includes the following:
  - Support Vectors: Identify a Brain tumor with high accuracy
  - Extreme Learning Machines: Help distinguish tumor types
  - K-Nearest Neighbors: For Tumor classification using precision and recall rates
  - Random Forest: How effective can we distinguish which tumor is present

## Methodology

**1. Data Wrangling. Preprocessing:**
- Handling Missing Data Values,
- Normalizing Pixel intensities,
- Performing Data augmentation for Cleansing

**2. Model Development and Training:**
- Were enforcing various models including: Decision Trees, SVMS, and Neural Networks
- Models extract features from each MRI images to form consensus on how a tumor is diagnosed, what tumor is classified as, and how can we treat it
- In addition, were portioning the data into training, validation, and testing sets

**3. Model Evaluation and Optimization:**
- We evaluate the performances of these models such as calculating Accuracy, Precision, Recalls, F-Scores, etc
- We want to ensure these models are effective in detecting a tumor as early as possible
- We also optimize models for better sensitivity and specificity

## Results

- The CNN (Convolutional Neural Network) model, built with multiple convolutional and pooling layers, significantly outperformed other methods including the SVM (Support Vector Machine) and Decision Trees.
- CNN achieved an accuracy of 92% and an excellent ROC AUC score of 0.967. Despite class imbalance (18,606 tumor vs. 3,066 normal images) identified tumors (precision: 98%, recall: 93%, F1-score: 95%), though it was less effective on normal images. The Decision Tree also achieved 92% accuracy but had a lower ROC AUC (0.775), indicating limited discrimination power.
- Linear SVM, while simpler, yielded lower overall accuracy (88%) and poorer performance on normal cases, highlighting its susceptibility to dataset imbalance. These findings underscore the superior diagnostic potential of CNNs in medical imaging for early, accurate, and automated detection of brain tumors.

```
Epoch 1/10
407/407 ━━━━━━━━━━━━━━━━━━━━ 355s 863ms/step - accuracy: 0.8588 - loss: 0.3994 - val_accuracy: 0.8238 - val_loss: 0.4217
Epoch 2/10
407/407 ━━━━━━━━━━━━━━━━━━━━ 357s 877ms/step - accuracy: 0.8713 - loss: 0.3125 - val_accuracy: 0.8055 - val_loss: 0.4586
Epoch 3/10
407/407 ━━━━━━━━━━━━━━━━━━━━ 342s 840ms/step - accuracy: 0.8878 - loss: 0.2694 - val_accuracy: 0.8987 - val_loss: 0.2265
Epoch 4/10
407/407 ━━━━━━━━━━━━━━━━━━━━ 342s 840ms/step - accuracy: 0.8965 - loss: 0.2368 - val_accuracy: 0.8265 - val_loss: 0.4093
Epoch 5/10
407/407 ━━━━━━━━━━━━━━━━━━━━ 409s 908ms/step - accuracy: 0.9076 - loss: 0.2130 - val_accuracy: 0.8881 - val_loss: 0.3327
Epoch 6/10
407/407 ━━━━━━━━━━━━━━━━━━━━ 342s 841ms/step - accuracy: 0.9161 - loss: 0.1929 - val_accuracy: 0.8955 - val_loss: 0.2843
Epoch 7/10
407/407 ━━━━━━━━━━━━━━━━━━━━ 338s 830ms/step - accuracy: 0.9213 - loss: 0.1845 - val_accuracy: 0.9170 - val_loss: 0.2103
Epoch 8/10
407/407 ━━━━━━━━━━━━━━━━━━━━ 336s 824ms/step - accuracy: 0.9207 - loss: 0.1794 - val_accuracy: 0.8851 - val_loss: 0.2989
Epoch 9/10
407/407 ━━━━━━━━━━━━━━━━━━━━ 409s 1s/step - accuracy: 0.9255 - loss: 0.1712 - val_accuracy: 0.9123 - val_loss: 0.2200
Epoch 10/10
407/407 ━━━━━━━━━━━━━━━━━━━━ 343s 843ms/step - accuracy: 0.9378 - loss: 0.1495 - val_accuracy: 0.9156 - val_loss: 0.2081
136/136 ━━━━━━━━━━━━━━━━━━━━ 30s 222ms/step - accuracy: 0.9224 - loss: 0.1943
Test Accuracy: 0.9217993021011353
```

Figure 1. CNN model accuracy and loss progression over epochs

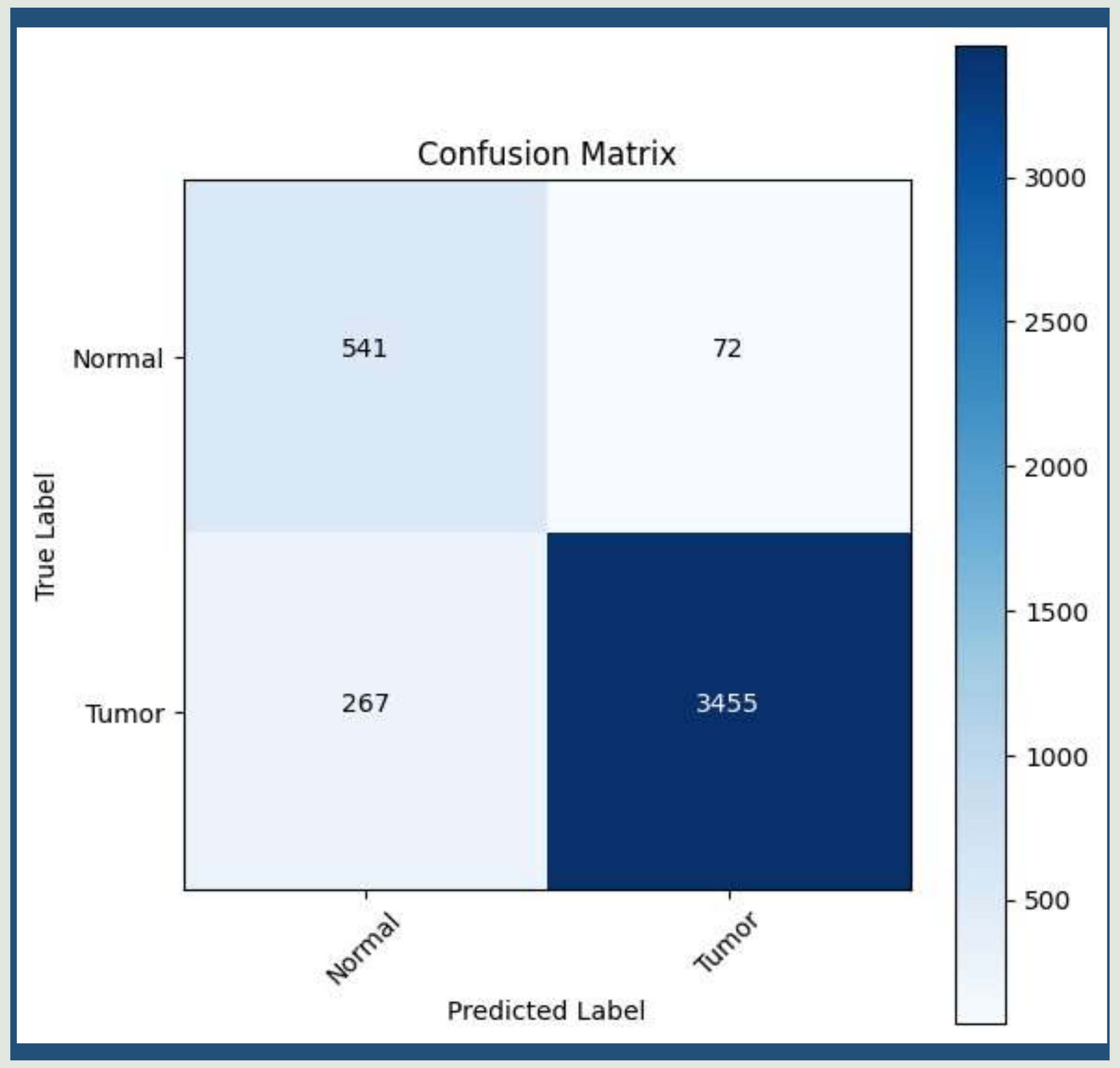


Figure 2. Confusion matrix for CNN predictions

## Conclusion

This study demonstrated the effectiveness of convolutional neural networks (CNN) for accurately detecting brain tumors from MRI images. The CNN achieved superior accuracy (92%) and outstanding discriminative ability (ROC AUC: 0.967), despite challenges posed by class imbalance. These findings highlight the potential of CNNs to significantly enhance early diagnosis, reduce human errors, and allow early personalized treatment planning. Continued studying and refinement of CNN models could substantially improve patient outcomes.

## Recommendations

- In this study, CNNs especially excel in detecting brain tumors from MRI images.
- Image pre-processing, data augmentation, and feature selection improve the accuracy
- The CNNs achieved high ROC AUC scores, indicating strong predictive performance.
- CNNs especially excel in detecting brain tumors from MRI images.
- Some limitations in CNNs remain, requiring further refinement.
- Ongoing collaboration between medical and data experts is vital to fully realize these technologies.

## Acknowledgements

We love to express of deepest gratitude to Mr Hossein Hashemi for supplying us with dataset with all forms of Tumor brain MRI Images. His dataset helped us in performing our Analysis and Results exactly how we expected.