PREDICTION OF COMPUTED TOMOGRAPHY FINDINGS USING SERUM METABOLITES MEASURED AFTER TRUMATIC BRAIN INJURY

1.0 BACKGROUND

Efficiently triaging traumatic brain injury (TBI) patients arriving at the Emergency Room (ER) greatly improves their management. The idea of having a point of care test kit that can predict TBI patients that are likely to have a positive or negative CT finding will shorten Arrival to Decision time (ADT), minimize unnecessary exposure to the ionizing radiation of a CT scan, and reduce the cost of management especially in regions where CT scans aren't readily available. On this note, a study done by Dickens et al measured serum metabolites that extravasates the blood brain barrier following TBI as a predictor of the severity of CT findings using the Marshall Grade [grade 1 (CT negative), grade 2–6 (CT positive)].

Marshall grade 2–4 is further categorized as (diffuse injury) while grade 5–6 (mass lesion). Obviously, depending on the extent of injury a TBI patient will require a CT scan. However, the Glasgow Comma Scale (GCS) which is used to clinically classify the severity is quite subjective and not so accurate. Hence, TBI patients that may have a positive CT finding may be under classified as mild.

2.0 OBJECTIVE

This study aims use Machine Learning (ML) classification models to determine whether:

1) The measured extravasated serum metabolites can effectively be used to differentiate specifically mild TBI patients with positive CT findings from those with negative findings.
2) Differentiate those with diffuse injury from those with mass lesion.
3) Compare model performance with that of previous study.

3.0 METHODOLOGY

3.1 Data Preprocessing

The shape of the original data set was (593, 414) and after cleaning, it was reshaped to (573, 392) by deleting rows that had too many missing values (especially values of the target feature) and columns that weren't necessary for prediction such as the patient identification number etc. However, all the measured metabolites were included in the independent features and the remaining missing values were replaced with zeros. The target features to be predicted were the CT Findings (positive or negative) and the Marshal grade (negative, diffuse injury or mass lesion).

3.2 Training of models

I selected the best model from the following Machine learning classification techniques; Logistic regression, Support Vector Machines (SVM), K Nearest Neighbours (KNN), Naive Bayes, Decision tress, Ensemble techniques (Bagging), Random Forest, Adaboost, and Xtreme Gradient boost.

Steps in training the models.

a) Prediction of CT findings (Positive or Negative)

- Import data.
- Checked for missing values.
- Univariate analysis of individual features in data set.
- Splitting of data into training and test sets
- Scaling of training set
- Where possible, I performed a grid search cv to obtain best parameters.
- Trained model on the best parameters and calculated the evaluation metrics; confusion matrices and ROC/AUC curve.

b) Prediction of Marshall grades (CT negative, Diffuse injury, Mass lesion) by repeating some of the steps mentioned above.

Data Source: Alex Dickens, PhD. Project Leader & Academy Research Fellow, Turku Bioscience Centre. Docent, Department of Life Technologies
August 2023

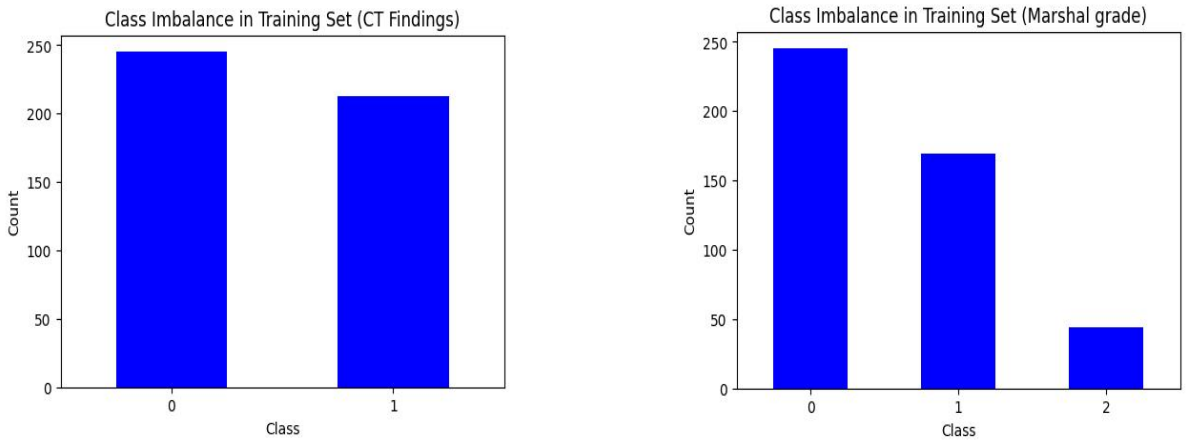
A. Models Evaluation Metrics in Predicting CT Findings as either Positive or Negative

Model	Train Accuracy	Test Accuracy	Precision	Sensitivity	Specificity	F1 Score	Confidence Interval
Logistic Regression*	0.79	0.79	0.78	0.78	0.80	0.78	0.5641, 0.6540
SVC*	0.75	0.76	0.72	0.81	0.71	0.76	0.5669, 0.6339
KNN*	0.70	0.72	0.73	0.65	0.78	0.69	0.5576, 0.6289
NB	0.68	0.70	0.66	0.78	0.63	0.71	0.6335, 0.6892
Decision Tree	0.73	0.75	0.72	0.78	0.73	0.75	0.6375, 0.7207
Random Forest	0.78	0.78	0.75	0.81	0.75	0.78	0.6577, 0.7422
Adaboost	0.78	0.72	0.70	0.70	0.73	0.70	0.6323, 0.7257
XGBoost	0.79	0.76	0.74	0.78	0.75	0.76	0.6578, 0.7526

B. Models Evaluation Metrics in Predicting Marshal Grades

Model	Train Accuracy	Test Accuracy	Precision	Sensitivity	Specificity	F1 Score	Confidence Interval
Logistic Regression	0.72	0.71	0.70	0.71	0.71	0.69	0.5152, 0.5909
SVM	0.84	0.56	0.59	0.56	0.56	0.57	0.3769, 0.4330
KNN	0.61	0.60	0.57	0.60	0.60	0.55	0.5075, 0.5987
NB	0.58	0.53	0.59	0.53	0.53	0.55	0.4981, 0.5768
Decision Tree	0.69	0.68	0.64	0.68	0.68	0.66	0.5789, 0.6567
Random Forest	0.74	0.68	0.70	0.68	0.68	0.69	0.5685, 0.6323
Adaboost	0.70	0.68	0.67	0.68	0.68	0.67	0.5725, 0.6282
XGBoost	0.80	0.67	0.66	0.67	0.67	0.66	0.5701, 0.6654

Table 1.0 KNN = K Nearest Neighbour Classifier, SVM = Support Vector Machine, NB = Naïve Bayes, XGBoost = Extreme Gradient Boost. Values are rounded up to the nearest 2 decimals.



A **B**

Fig 1.0 Imbalance in the training data sets. **A** CT negative as 0, CT positive as 1. **B** CT negative as 0, Diffuse lesion as 1, Mass lesion as 2.

Data Source: Alex Dickens, PhD. Project Leader & Academy Research Fellow, Turku Bioscience Centre. Docent, Department of Life Technologies
August 2023

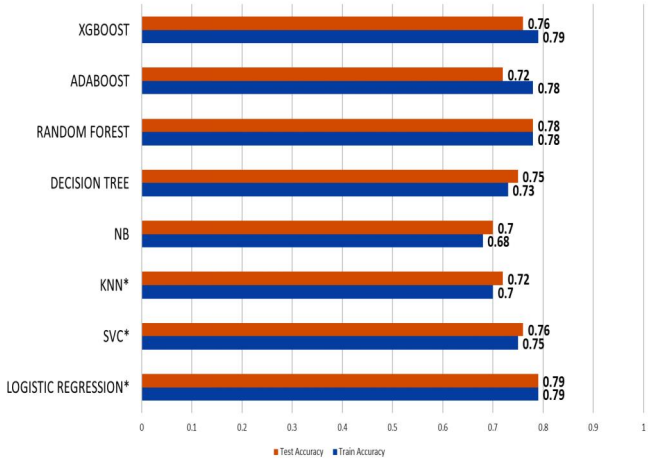


Fig 2.0 Models Train and Test Accuracies in Predicting CT Findings

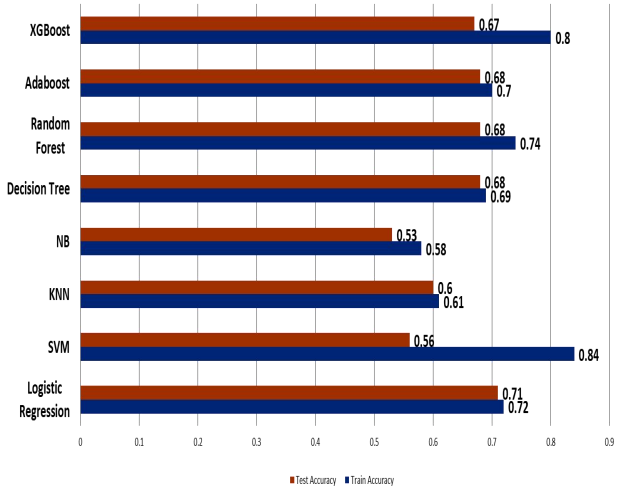


Fig 3.0 Models Train and Test Accuracies in Predicting Marshal Grade

4.0 RESULTS

4.1 Evaluation metrics

In the previous study, the prediction model used was logistic regression which had an AUC of 0.77 for the discovery cohort (Turku) and 0.73 for the validation cohort (Cambridge) when predicting if CT findings will be positive or negative. However, in differentiating between a diffuse or a mass lesion the AUC for the discovery cohort was 0.87 while that of the validation cohort was 0.68. In this study, the data used to build the models was for the discovery cohort. Logistic regression and SVM classifier both had an AUC of 0.85 and 0.87 respectively when predicting if CT findings will be positive or negative **(Table 1.0)**. However, using the mean test accuracy, XGBoost, Adaboost and Random forest had the highest accuracy of 0.70, 0.67, and 0.69 respectively. In differentiating the marshal grades, XGBoost, Decision tree and KNN had the highest mean test accuracy of 0.61.

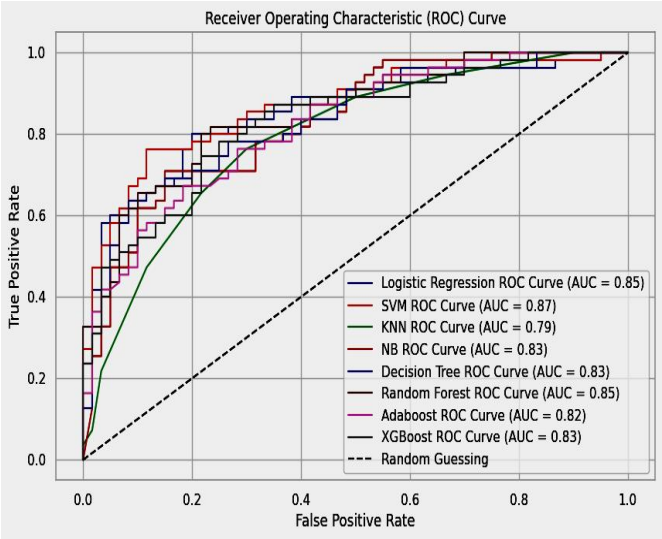


Fig 4.0 ROC/AUC of the models in predicting CT Findings

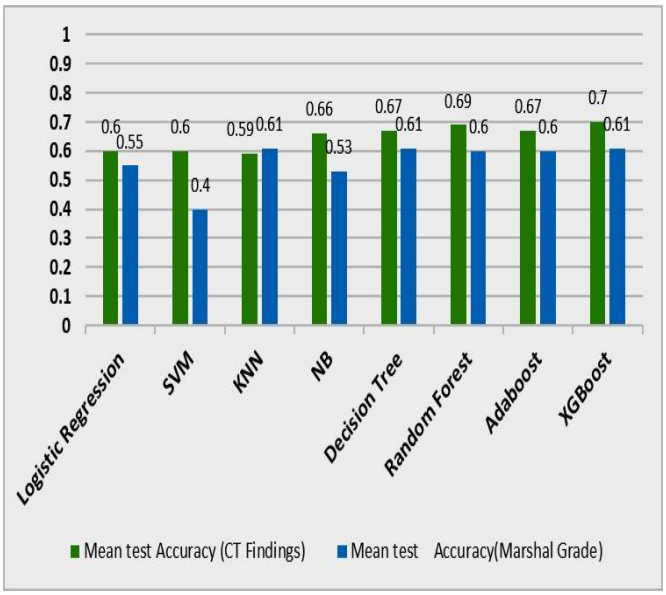


Fig 5.0 Mean test accuracies in predicting CT Findings and Marshal Grades after a 10 k-fold cross-validation. Models are better at predicting CT Findings than Marshal Grades.

Data Source: Alex Dickens, PhD. Project Leader & Academy Research Fellow, Turku Bioscience Centre. Docent, Department of Life Technologies
August 2023

4.2 Models performance in predicting positive or negative CT Findings

The mean test accuracy ranged from 0.59 to 0.70. Here we present the confusion matrix of the 3 best models (XGBoost, Adaboost and Random forest) in terms of their mean test accuracy

4.2.1 XGBoost Classifier

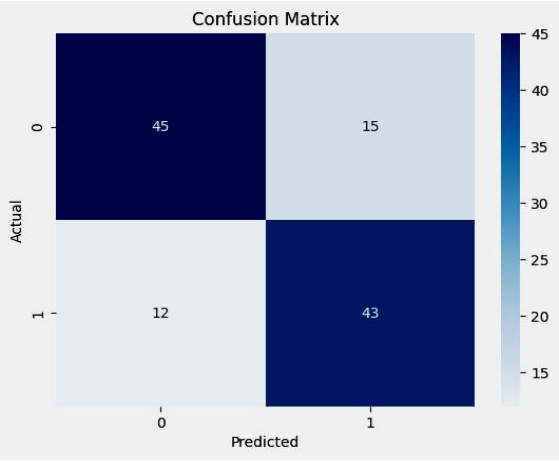


Fig 6.0 Confusion matrix for the XGBoost classifier. Positive CT findings as 1, Negative as 0.

Prediction on the test data set that was split from the training set. Model made a total of 57 predictions for negative findings out of which 12 were wrongly classified as positive. Similarly, a total of 58 predictions was made for positive findings out which 15 were wrongly classified as negative.

4.2.2 Adaboost Model

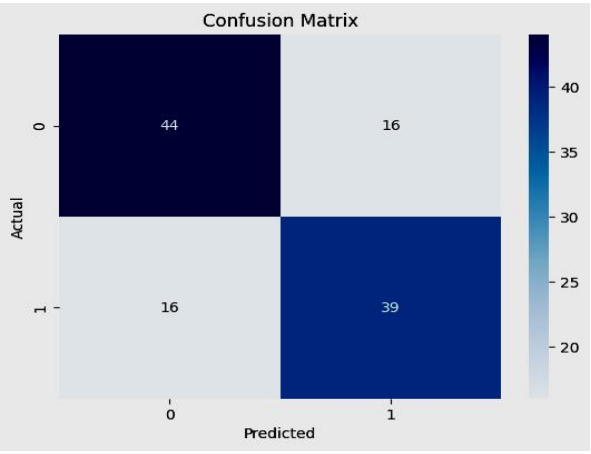


Fig 8.0 Confusion matrix for Adaboost. Positive CT findings as 1, Negative as 0.

Model made a total of 60 predictions for negative findings out of which 16 were wrongly classified as positive. Also, a total of 55 predictions was made for positive findings out of which 16 were wrongly classified as negative.

4.2.3 Random Forest Model

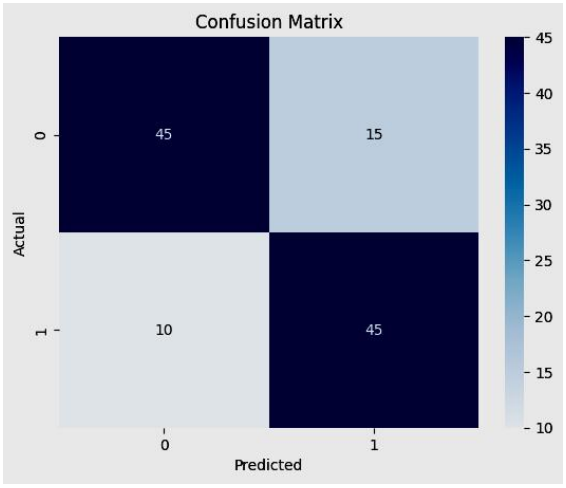
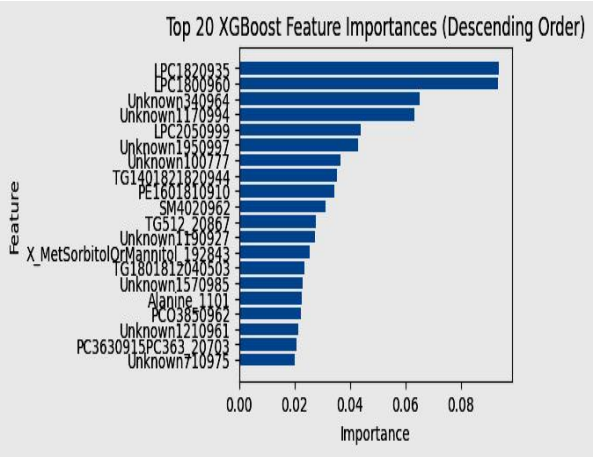


Fig 7.0 Confusion matrix for Random Forest. Positive CT findings as 1, Negative as 0.

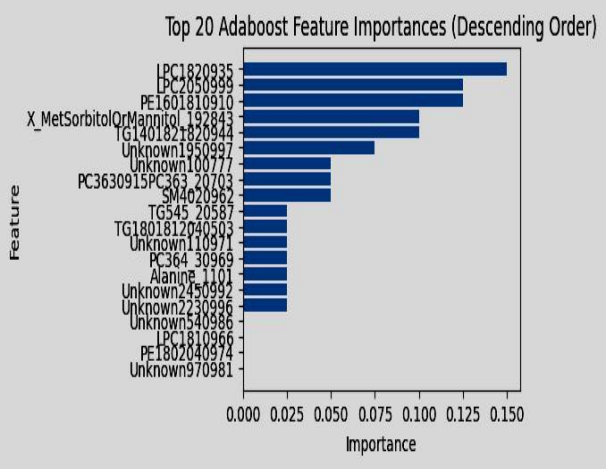
In predicting the test data, model made a total of 55 predictions for negative findings out of which 10 were wrongly classified as positive while a total of 60 predictions was made for positive findings out which 15 were wrongly classified as negative.

Data Source: Alex Dickens, PhD. Project Leader & Academy Research Fellow, Turku Bioscience Centre. Docent, Department of Life Technologies
August 2023

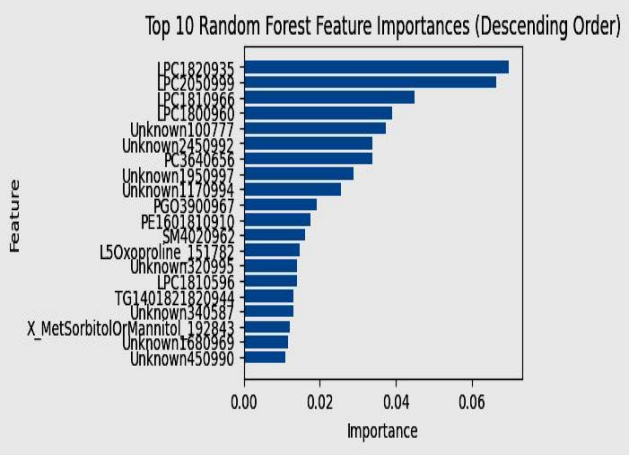
4.2.4 Feature Importance influencing CT Findings



A



B



C

Fig 9.0 10 most significant metabolites that determines positive or negative CT findings. **A** XGBoost, **B** Adaboost and **C** Random Forest**. 'LPC1820935', 'LPC2050999', 'Unknown1950997',** and **'TG1401821820944'** appears to be the most important determinants as they featured in all 3 models feature importance**.**

4.2.5 Distribution within Significant determinants of CT findings

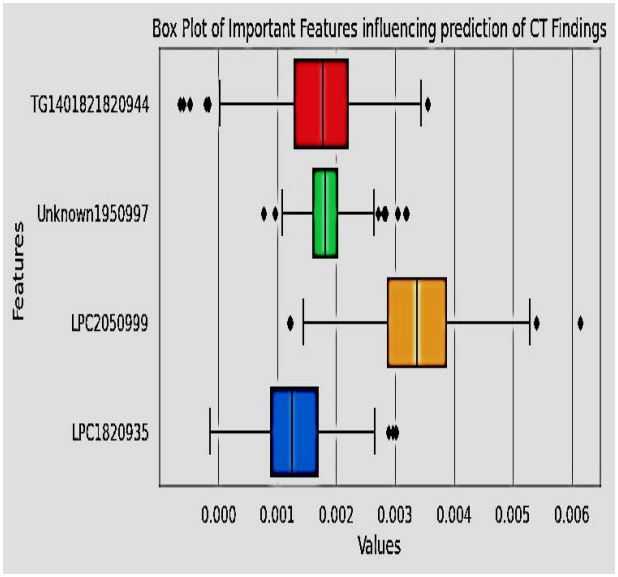


Fig 10.0 Boxplot and whiskers of the distribution of 4 significant determinants of positive or negative CT findings. Outliers are identified with asterix.

4.3 Models performance in predicting Marshal grades (Negative, Diffuse, Mass lesion)

The mean test accuracy ranged from 0.4 to 0.61. Similarly, we present the confusion matrix of the 3 best models (XGBoost, Decision Tree and KNN) in terms of their mean test accuracy.

. 4.3.1 XGBoost Classifier

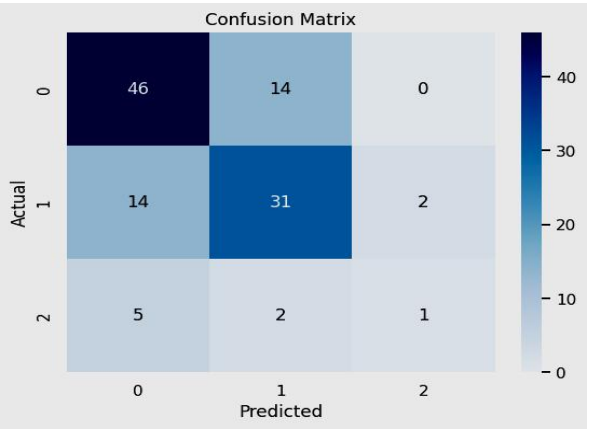


Fig 11.0 Confusion matrix for XGBoost Classifier. Mass lesion as 2, Diffuse injury as 1, Negative as 0.

In predicting the test data, model made a total of 65 predictions for negative findings out of which 19 were wrongly classified as diffuse injury, a total of 47 predictions was made for diffuse injury out of which 16 were wrongly classified as negative, and a total of 3 predictions was made for mass lesions out which 2 were wrongly classified as diffuse injury and non as negative finding.

4.3.2 Random Forest

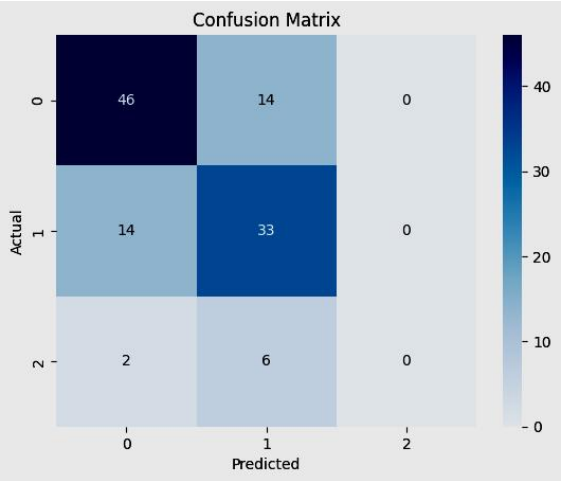


Fig 12.0 Confusion matrix for Decision Tree. Mass lesion as 2, Diffuse injury as 1, Negative as 0.

In predicting the test data, this model made a total of 62 predictions for negative findings out of which 14 were wrongly classified as diffuse injury and 2 as mass lesions, a total of 53 predictions was made for diffuse injury out of which 14 were wrongly classified as negative, and 6 as mass lesions while no predictions was made for mass lesions.

4.3.4 KNN

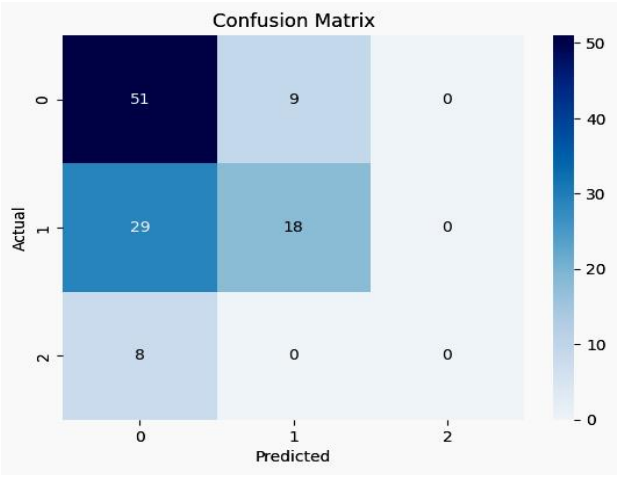
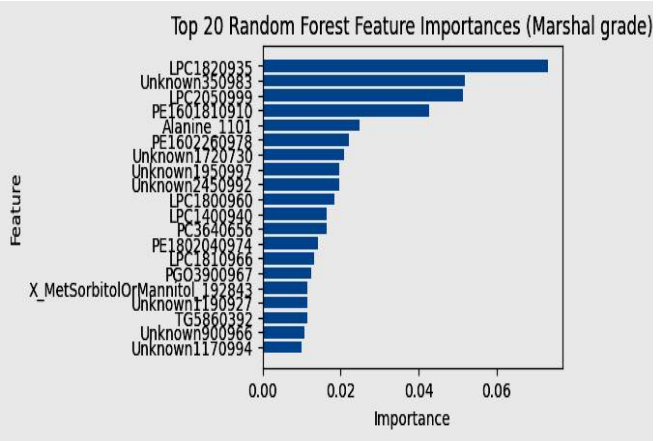


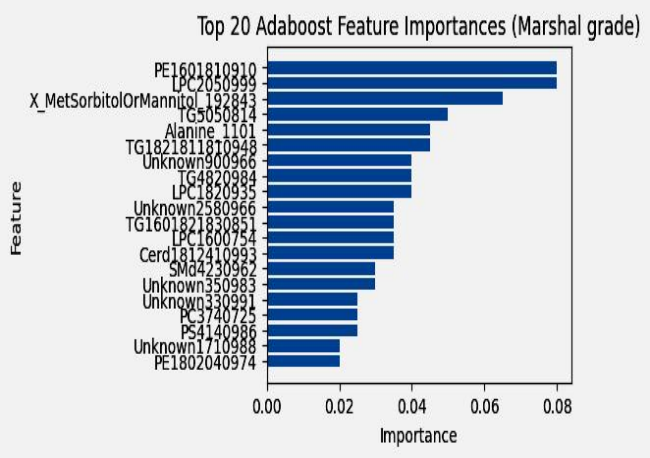
Fig 13.0 Confusion matrix for KNN Classifier. Mass lesion as 2, Diffuse injury as 1, Negative as 0.

In predicting the test data, model made a total of 88 predictions for negative findings out of which 29 were wrongly classified as diffuse injury and 8 as mass lesion. A total of 33 predictions was made for diffuse injury out of which 9 were wrongly classified as negative and non as mass lesion, while no predictions was made for mass lesions.

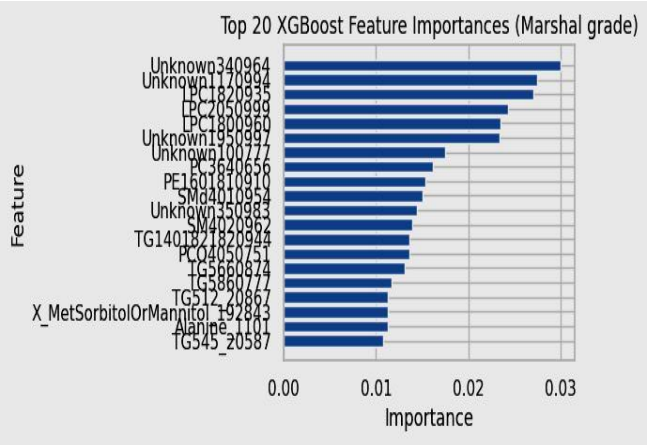
4.3.5 Feature Importance influencing prediction of Marshal grades.



A



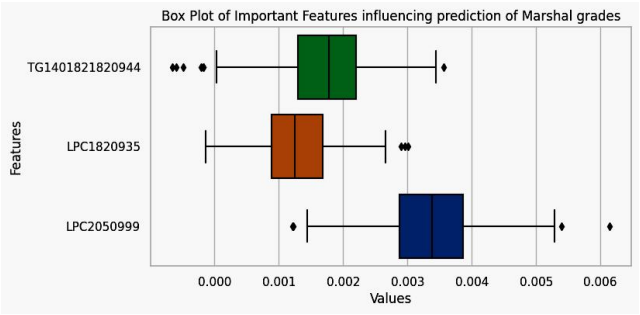
B



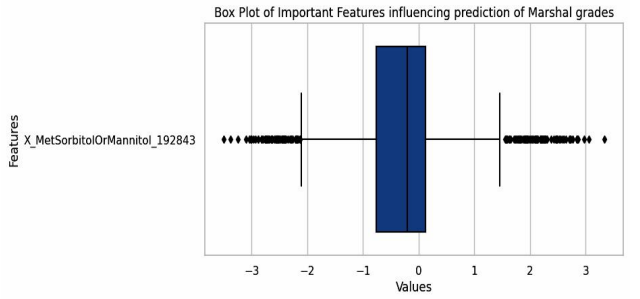
C

Fig 14.0 10 most significant metabolites that determines Marshal grades (negative, diffuse injury, and mass lesion). **A** Random Forest, **B** Adaboost and **C** XGBoost. **'LPC1820935','LPC2050999', 'TG1401821820944', and X_MetSorbitolOrMannitol_192843** appears to be the most important determinants as they featured in all 3 models feature importance.

4.3.6 Distribution within Significant determinants of CT findings



A



B

Fig 15.0 Boxplot and whiskers of the distribution of 4 significant determinants of Marshal grades (negative, diffuse injury, and mass lesion). Outliers are identified with asterix.

5.0 DISCUSSION

5.1 Models performance

Overall, the models support the hypothesis that beyond random guessing, serum metabolites measured following a TBI can be used as a predictor of CT findings. This is because, most of the models had an accuracy of at least 0.60 with minimal overfitting suggesting that it should do well in predicting unseen data. The range of prediction accuracy on the test data for positive or negative CT findings was between 0.68 to 0.79 however, XGBoost classifier had the highest mean test prediction accuracy of 0.70 while KNN classifier had the least mean accuracy with 0.59. In predicting the marshal grade (negative, diffuse

injury, and mass lesion), the range of prediction accuracy was between 0.53 to 0.71 with XGBoost, Decision Tree, and KNN classifier having the highest mean test accuracy of 0.61 while SVM had the least mean test accuracy of 0.4. **(Fig 5.0).** Other evaluation metrics such as the sensitivity, specificity, precision and F1 score were comparable to the test prediction accuracy.

In comparison to the model performance in the previous study; 0.77 accuracy prediction for positive or negative finding, 0.87 accuracy for mass lesion or diffuse injury in the discovery cohort, the XGBoost classifier in this study has a better prediction accuracy for CT Findngs but not for Marshal Grades.

5.2 Feature importance

Using XGBoost, Adaboost and Random-forest techniques to obtain feature importance, the metabolites by the names **'LPC1820935', 'LPC2050999', 'Unknown1950997',** and **'TG1401821820944'** seems to be of utmost importance in determining positive or negative CT findings because they were present in all the 3 models feature importance **(Fig 7.0 and 8.0).** In the previous study, 3 metabolites were used to train the logistic regression model which were: an unknown phenolic compound, isovaleryl glucuronide and 2-hydroxybutyric acid. Since the metabolites identified by this study's model feature importance are unnamed, we can't confirm if they were the ones used in the previous study. Similarly, the metabolites by the names **'LPC1820935','LPC2050999', 'TG1401821820944', and X_MetSorbitolOrMannitol_192843** were identified by XGBoost, Adaboost and Random-forest techniques **(Fig 12.0 and 13.0)** as some of the most significant in determining diffuse injury or mass lesion. In this regard, the previous study identified 6 metabolites: 2-Aminobutyric acid, Amino acid, Acetoacetic acid, Pentitol, 3-desoxy, Inositol, and Ribonic acid. Similarly, they can't be confirmed as they are unnamed as well.

6.0 CONCLUSION

In conclusion, serum metabolites measured following traumatic brain injury can be used as a predictor of CT findings, more so in distinguishing diffuse injury from mass lesion. This is confirmed in this study because, the least prediction accuracy by a model was 0.60 while the XGBoost classifier had the highest mean test prediction accuracy on the discovery cohort of 0.70 for predicting positive or negative CT findings and 0.61 for predicting diffuse injury or mass lesion. Furthermore, in comparison to the previous study the XGBoost classifier in this study has a better prediction accuracy on the discovery cohort for CT Findings. The metabolites identified in this study to have the most significant impact on prediction could not be corroborated with that used in the previous study for model training as they were unnamed.

7.0 LIMITATIONS

- Due to my computer's low processing power and RAM, it was very time consuming doing a grid search cross validation to fine-tune hyperparameters to get the best possible combinations for use in training my model. Hence, I manually adjusted the hyperparameters for most of the models.
- I assumed all the columns I used in training the models were all metabolites with no proteins. This is because some columns were unnamed.
- I assumed the data set I got was that of the discovery cohort.

8.0 SUGGESTIONS FOR IMPROVING PREDICTION ACCURACY

- Depending on how the model is deployed, adding the individual components of the mild GCS score (Best eye, motor, and verbal response) in training the models may increase prediction accuracy. However, this may not be suitable in an emergency setting where quick decisions are to be made.
- Perform further data preprocessing and fine tuning of hyperparameters using

the important metabolites identified in this study to improve model accuracy.

9.0 REFERENCES

1. Thomas I, Dickens AM, Posti JP, Czeiter E, Duberg D, Sinioja T, Kråkström M, Retel Helmrich IRA, Wang KKW, Maas AIR, Steyerberg EW, Menon DK, Tenovuo O, Hyötyläinen T, Büki A, Orešič M; CENTER-TBI Participants and Investigators. Serum metabolome associated with severity of acute traumatic brain injury. Nat Commun. 2022 May 10;13(1):2545. doi: 10.1038/s41467-022- 30227-5. PMID: 35538079; PMCID: PMC9090763.

2. Dickens AM, Posti JP, Takala RSK, Ala-Seppälä H, Mattila I, Coles JP, Frantzén J, Hutchinson PJ, Katila AJ, Kyllönen A, Maanpää HR, Newcombe V, Outtrim J, Tallus J, Carpenter KLH, Menon DK, Hyötyläinen T, Tenovuo O, Orešic M. Serum Metabolites Associated with Computed Tomography Findings after Traumatic Brain Injury. J Neurotrauma. 2018 Nov 15;35(22):2673-2683. doi: 10.1089/neu.2017.5272. Epub 2018 Aug 21. PMID: 29947291.

3. Reith FC, Synnot A, van den Brande R, Gruen RL, Maas AI. Factors Influencing the Reliability of the Glasgow Coma Scale: A Systematic Review. Neurosurgery. 2017 Jun 1;80(6):829-839. doi: 10.1093/neuros/nyw178. PMID: 28327922

Data Source: Alex Dickens, PhD. Project Leader & Academy Research Fellow, Turku Bioscience Centre. Docent, Department of Life Technologies
August 2023