Application of Machine Learning Models to Predicting Brain Hemorrhages

Bhargav Parsi (804945591), Kevin Chiu (704463286)

***Abstract*— In this paper, we apply different machine learning models to previous work done by Dr. Yannan Yu at UCLA on the prediction of hemorrhagic transformations (HT) in acute ischemic stroke (AIS) patients and report on our results.**

I. Data Visualization

In this section, we examine the techniques used to visualize the dataset. This provides a sense as to how our data appears and hence, the motivation behind our approach to optimizing the prediction of HT development. We fist examined the distribution of patients in this dataset, which can be seen in Fig. 1. Observation shows that patient data samples for each patient are not distributed uniformly. In Fig. 2, we can see that the labels in the dataset, HT or non-HT, are distributed uniformly. From this, we can conclude that there is no need to subsample our data to train our model.

In each 3 x 3 patch in the MRI scan of the brain, the following 4 quantities were extracted:

- DWIb0
- DWIb1000
- PWI
- AIF

Diffusion-weighted imaging (DWI) is a form of MR imaging based upon estimates of the rate of water diffusion at a location. Using these measurements, contrast can be developed in an image and these water molecule diffusion patterns can show microscopic details about tissue. The b values associated with DWI measure the degree of diffusion weighting applied. A larger b value is utilized to measure slower moving water molecules and smaller diffusion distances when compared to smaller b values. Water mobility is highly dependent on the cellular environment and as a result, DWI has been found to be very sensitive to the early changes an onset of a stroke brings to the body [9].

Perfusion weighted imaging (PWI) in MRI often uses dynamic susceptibility contrast (DSC) MRI, which takes T2*-weighted gradient echo pulse sequences, for perfusion scanning. This data can be processed to provide information on the status of the tissue and detect impaired perfusion in the ischemic core and surrounding brain areas. PWI combined with DWI can therefore be a powerful tool for determining potentially salvageable tissue in AIS patients.

MRI is powerful imaging method of assessing cerebral perfusion and can provide cerebral blood flow measurements, and central to this quantification is the arterial input function (AIF), which describes the time-dependent contrast concentration input to the target tissue. Central artifacts in cerebral blood flow quantification, such as bolus delay and peak truncation effects, are directly linked to the AIF.

II. Proposed methods

In this section we will define the various machine learning models we have used for this project. It is essential to use complex models as we can see that the number of features can be significant enough that a simple model might not work well. Also, since the dataset consists of 2 different types of features, time dependent and non time dependent, we decided to analyze them separately with different models. We train different models with different training sets, but on the same test set. The training set and test set are chosen such that the intersection of the patient IDs in each of the set is a null set, which means that the model is trained on a completely different set of patients and tested on a set with entirely new patient data, ensuring the model is not biased.

We first analyze the data using a crude model, which we denote as simple features selection, which takes in all the features at once and makes a prediction. Second, we use only the features which are non-time dependent, which we denote as non-time series feature selection. Third, we make use of deep learning to make a model, which can do the prediction using only the time dependent features. Finally, we make use of a multi input model, which takes in both type of features to make the prediction.

A. Simple Feature Selection

1) Decision Tree Model: In this model we have used a decision tree to perform our classification task. We use the Scikit Learn package [5] in Python to implement the model. This model learns simple decision rules, which can be inferred from the data.

We have used the ”gini” split criterion in the decision tree we have implemented. The number of minimum samples required for a split is taken as 2. We have considered all the features for the split while all the other parameters of the decision tree are taken to be default.

The merits of using this model include the following:

- Simple to understand
- The cost of using the tree is logarithmic in the number of data points used to train the tree.
- Uses a white box model. By contrast, in a black box model, results may be more difficult to interpret.

Some problems with this approach include the following:

- It has very high probability of over-fitting the data due to a tendency to create overly complex trees.
- It can be very unstable and small variations in data might lead to a completely different tree.

2) XGBoost Model: In this model, we used an open source version of the model presented in [6]. This model uses an ensemble of decision trees instead of a single decision tree. The prediction of this model sums of predictions of all the trees together. This has an inbuilt regularization module which takes care of the regularization unlike a simple decision tree. The goal of this library is to push the extremes of the computation limits of machines to provide a scalable, portable and accurate library. Only the default parameters provided in the tool are utilized.

B. Non-time series feature selection

Here, we have selected only the DWI features which are time independent. In the previous models, boths time-dependent and time independent features are used together, which might lead to inaccurate decisions because the model treats all of the features similarly when they are distinct. This drove the motivation behind using only the non-time series feature selection method.

We have similarly used the decision tree and XGBoost using the same set of parameters, but with a decrease in the feature set size from 618 to only 18 features, which are non-time series features. We observed that the time taken for the model to train is much lower than the model which uses all of the features, which is expected due to the reduced number of features.

C. Time series feature selection

Here we consider the remaining 600 features which come from the PWI and AIF. There are 9 types of PWI and 1 type of AIF given in the sample data. All have values in 60 time steps. Therefore, each type of PWI and AIF for 1 patient can be represented as a vector of numerical values in which each numerical value is associated to 1 time step. The Long Short Term Memory algorithm, a variation of a simple recurrent neural network (RNN), suits very well for this purpose, as previous work shows it has strong performance in time series analysis [7]. A major motivation to opting for the LSTM over the RNN is that a LSTM doesn't have the vanishing gradient problem that a simple RNN would exhibit. The number of outputs of an LSTM is the number of LSTM cells used so a fully connected layer after the LSTM was used to convert the number of outputs to 2, the number of classes.

We have used the Keras library [8] in Python using a TensorFlow backend with GPU to make our models. Using Keras we can make our own customizable model within a few lines of code.

D. Multi Input Model

This model takes in both types of features, time-dependent and time independent, simultaneously, but analyses each of them separately. We use a LSTM similar to the one in the previous method to generate the probabilities as outputs and we concatenate them with the rest of the 18 non-time dependent features. A fully connected layer is inserted to reduce the dimension to 2, the number of classes. As expected, this model takes much longer to train as the number of parameters is higher compared to our previous methods. Also, another consideration is that there is a high chance of over-fitting as the number of trainable parameters are comparable with the amount of data we have. As a result, there is a limit to the complexity the model can have as the number of parameters would shoot up and result in more than the amount of data we have.

III. Results

A. Evaluation Metrics

For our evaluation metric, area under the curve along with a minimum of a 95% confidence interval is used. This is the same metric utilized in work done by Dr. Yu in [2], which we will use as a baseline to compare our models against. We have used the bootstrapping technique where we run around 100 iterations and randomly sample a set from the test set and interpret the AUC of that test set. All 100 AUCs are sorted and the first 2.5% and the last 2.5% are discarded. We then report the minimum and maximum values from the set, which is the confidence interval.

From the Figures section, we can observe that as the training data size increases, the AUC increases across all models, which is expected behavior. Comparisons across models show the decision tree performs poorly relative to other models. Of note is that the gradient boosting model using XGBoost outperforms previous work done in [2] even with a small number of training samples. If we examine the performance of the simple LSTM model, the best AUC achieved is approximately 0.82. However, by combining the simple LSTM model with a dense model, the AUC increases to 0.85, which shows improvement from Dr. Yu's work of 0.837. This supports our idea that by making use of more features and an increased complexity model, a more optimized decision could be made. All the LSTM models have been trained for 20 epochs and used the binary cross entropy as the loss function. We have used Adam optimizer for the optimization.

B. Visualization

We test our model on real life brain images and predict the output at every 3x3 patch of the image. We plot the heatmaps with the probability of the pixel that there would be a HT. Red indicates high probability and blue indicates low probability. These results are displayed in the Figures section below.

IV. CONCLUSIONS

Comparison of our results with the previous model developed by Dr. Yu in [1] shows our model demonstrates improved accuracy by at least 2%. Further improvements to the demonstrated model can be made with additional data and hyper-parameter optimization. Future work can be done by using an ensemble of different models and give a combined output. Continued improvements for HT prediction can lead to more accurate insight for medical professionals in assessing the potential risks.

V. FIGURES

A. *Data Visualization*

The following figure correspond to the data visualization report section, which examined the patient distribution versus samples collected, the number of samples that were HT and non-HT patients, and typical PWI response for a HT patient.

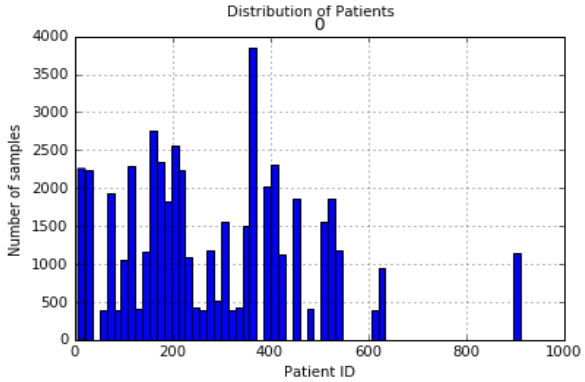

Fig. 1. Number of samples taken from each patient identified by an ID number.

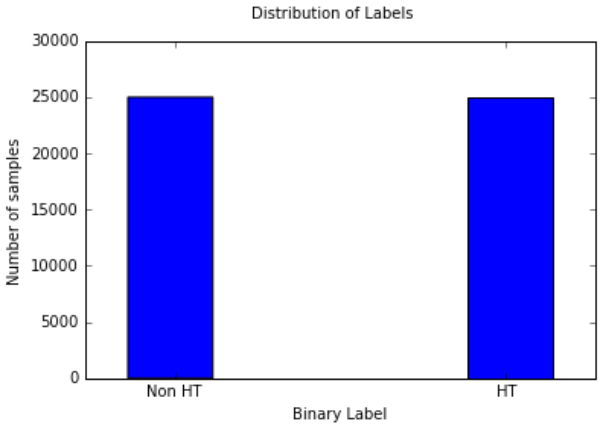

Fig. 2. Number of non-HT and HT samples is split equally at 25,000 apiece out of the 50,000 total samples.

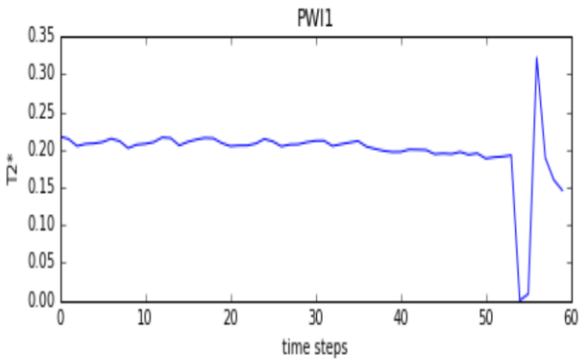

Fig. 3. Typical PWI T2 response for a HT patient across 60 time steps.

B. *Results for Various Models*

The following figures correspond to the results reported.

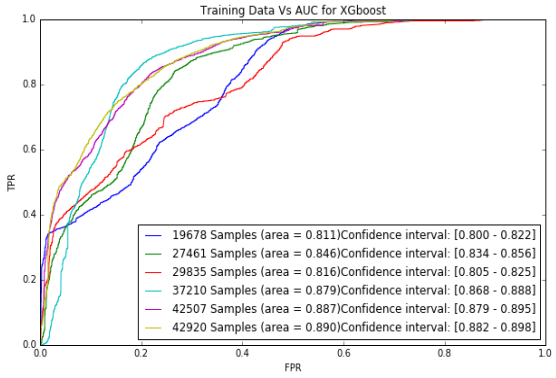

Fig. 4. Results for model using varying sample sizes with the XGBoost model with all features included.

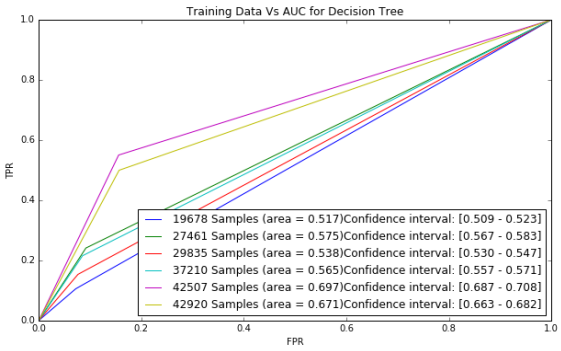

Fig. 5. Results for model using varying sample sizes with the decision tree model with all features included.

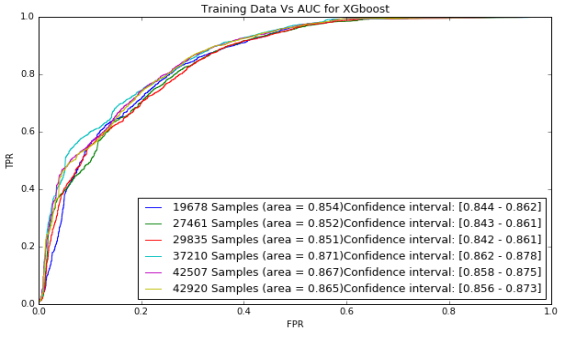

Fig. 6. Results for model using varying sample sizes with the XGBoost model with 18 features (non time-series) included.

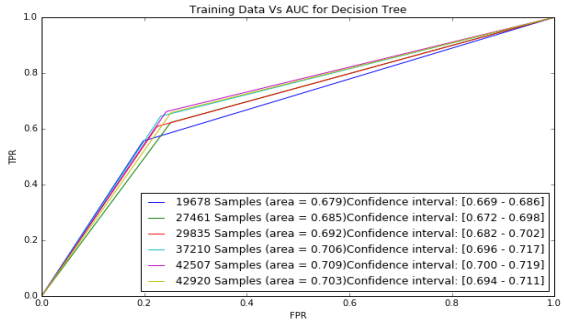

Fig. 7. Results for model using varying sample sizes with the decision tree model with 18 features (non time-series) included.

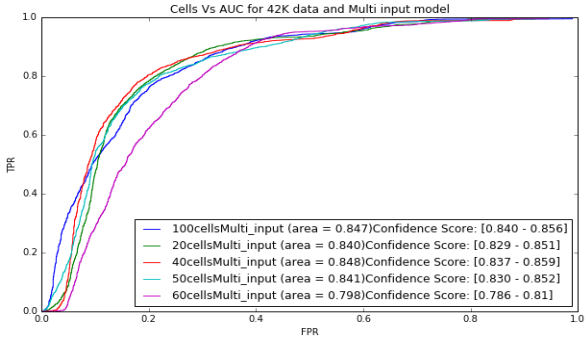

Fig. 8. Results for model using varying number of cells for the multi-input model.

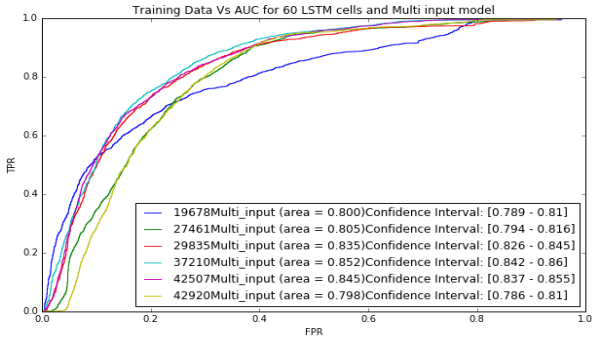

Fig. 9. Results for model using varying sample sizes with the multi-input model using 60 LSTM cells.

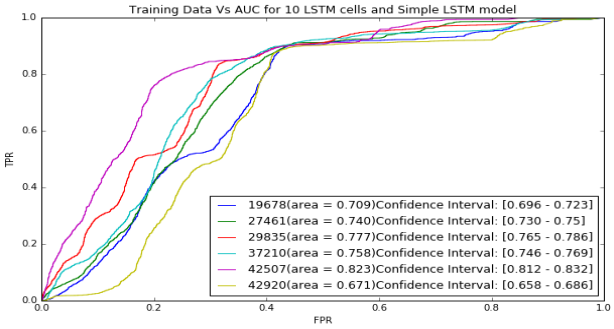

Fig. 10. Results for model using varying sample sizes with the simple LSTM model using 10 LSTM cells.

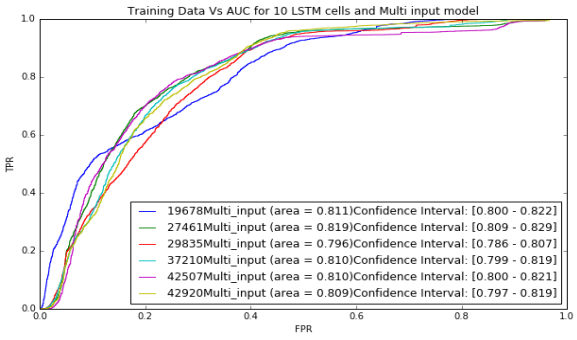

Fig. 11. Results for model using varying sample sizes with the multi-input model using 10 LSTM cells.

C. Generated Heatmaps

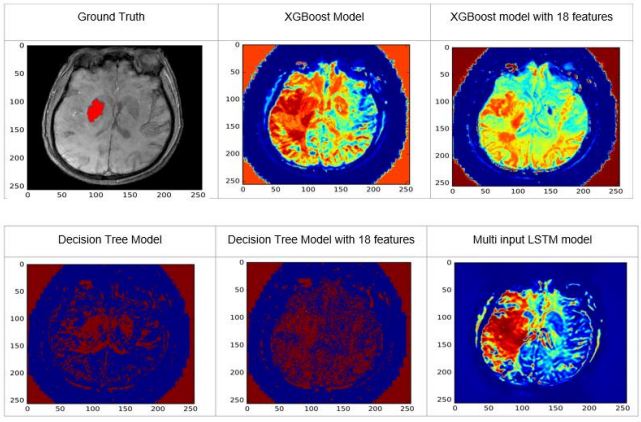

Fig. 12. Generated heatmaps for a sample patient with HT using the various models presented. Red areas indicate a high probability of bleeding while blue indicates a low probability.

REFERENCES

[1] U. Bashir, Diffusion weighted imaging — Radiology Reference Article, Radiopaedia.org. [Online]. Available: https://radiopaedia.org/articles/diffusion-weighted-imaging-1. [Accessed: 05-Jun-2018].

[2] Y. Yu, D. Guo, M. Lou, D. S. Liebeskind, and F. Scalzo. "Prediction of Hemorrhagic Transformation Severity in Acute Stroke from Source Perfusion MRI," *IEEE Transactions on Biomedical Engineering*, December, 2017. [Online]. Available: https://ieeexplore.ieee.org/document/8231213/. [Accessed: May 2, 2018].

[3] F. Scalzo, J. Alger, X. Hu, J. Saver, K. Dani, K. Muir, A. Demchuk, S. Coutts, M. Luby, S. Warach, and D. Liebeskind (2018). "Multi-center prediction of hemorrhagic transformation in acute ischemic stroke using permeability imaging features," *Magnetic Resonance Imaging*. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/23587928. [Accessed: April 30, 2018].

[4] Asadi, H., Dowling, R., Yan, B. and Mitchell, P. (2018). "Machine Learning for Outcome Prediction of Acute Ischemic Stroke Post Intra-Arterial Therapy," *PLOS*. [Online].Available:https://journals.plos.org/plosone/articleid=10.1371/journal.pone.0088225. [Accessed: May 2, 2018].

[5] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." Journal of machine learning research 12.Oct (2011): 2825-2830.

[6] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. ACM, 2016.

[7] Hochreiter, Sepp, and Jrgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.

[8] Chollet, Franois. "Keras." (2015): 128.

[9] V. J. Wedeen, P. Hagmann, W. Y. Tseng, T. G. Reese, and R. M. Weisskoff, Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging., Advances in pediatrics., Dec-2005. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/16247738. [Accessed: 06-Jun-2018].