Tytonidae Tympanometry

Applying Machine Learning to predict hearing loss using wideband tympanometry

Di Yao
22795234

Karan Rebello
22868277

Anitha Raghupathy
22773933

Cheng Nian
23053313

Aminul Islam
22884375

Daniel Chegwidden
21282744

October 2021

https://github.com/danielchegwidden/tytonidae-tympanometry

Contents

1 Tytonidae Tympanometry

The Tytonidae Tympaometry (TyTy) project was born out of a need for audiologists to better understand wideband tympanometry (WBT). WBT data is used to classify ears as either normal, or with conductive conditions such as Otitis Media. The subjects in this data are children, and not identifying conductive conditions has consequences throughout their life.

The societal cost of these conditions is estimated at $20,000 across 10 years as children struggle to learn and engage at school, amongst other effects [1]. This makes the identification and treatment of conductive conditions financially sound due to the average cost of each surgery being $5,000.

TyTy attempts to use WBT data to improve the classification accuracy of the current receiver operating characteristic (ROC) analysis that is performed, as well as provide audiologists with insights as to how to better analyse the large quantities of data that is being generated. If this is achieved, then clinicians are able to spend less time analysing data, and more time interacting with patients, resulting in more patients being seen and better care being delivered.

2 Models. Models. Models.

The Machine Learning approach that attempted to beat the ROC results involved Logistic Regression, Support Vector Machines, Decision Trees, Random Forests, and K-Nearest Neighbour as the modelling approaches taken. These were selected as having good classification applications, as well as a variety of complexity and interpretability implications.

The data was transformed through processing pipelines to ensure consistency and reproduceability. For each model, the appropriate pipeline could be used with different parameters to see how these transformations affected the final results. An example of this is pressure matching, where the full data was filtered down to a single observation per ear based on the pressure reading when the data was collected.

Logistic Regression was chosen as the base model and the starting point due to its simplicity and interpretability. This model creates a formula that outputs either a positive or negative class depending on the cutoff probability. For all of the different pipeline transformations, this model gave a baseline to beat using the same data. The ROC results that are currently being used may have slightly different transformations and it was important to be comparing like-for-like results. Tuning of the model hyperparameters resulted in 85% accuracy on the unseen test set, and this was what the more complex models attempt to beat.

One step up in complexity from Logistic Regression is Support Vector Machines (SVM). This model attempts to find a decision boundary, or hyperplane, that separates the data points into the positive and negative class. SVM works well with small and clean data which is what the WBT data is like, and can deal with the many frequencies present whilst maintaining its interpretability. The best results from the SVM model was 89% accuracy on the test set.

The next modelling approach explored was Decision Trees, which identify important features and the best split of the data at each of these features. This results in a binary decision at each node of the tree that splits into two leaf nodes where the process is repeated. The results of the training process show this in Figure 1 where the blue nodes how a positive class prediction, and the orange nodes show a negative one. The Decision Tree model is interpretable as the tree can be followed from the root node down to a specific leaf node to identify which features are contributing to the prediction. The colouring also indicates the confidence of the prediction with the darker colours indicating the degree of pureness in the node. The best results from the Decision Tree was 90% on the test set, which jumps up to 93% when using cross-validation.

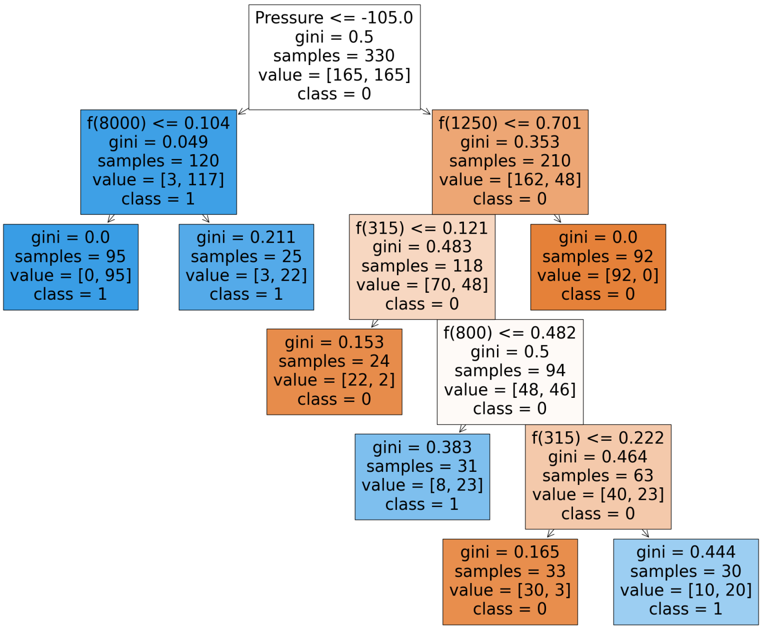

Figure 1: The best Decision Tree model after being trained

Building on from Decision Trees, Random Forests are an ensemble of many of these trees grouped together and then averaging out their individual results. This takes longer to run than a single tree, but uses a wisdom-of-the-crowd approach to reduce the variance from the forest. The downside of this approach is that the interpretablity drops due to the averaging across many trees, which means that the results or a specific tree are both inaccessible and irrelevant to the performance of the overall model. The performance of the Random Forest on the test set was 89%.

The final model considered was K-Nearest Neighbours, which is an unsupervised technique that attempts to group similar data points together. The number of groups that this model uses is the hyperparameter k, which results showed to be 116 in the optimal model, leading to an accuracy of 89%. However this does not make much sense as ideally k would be 2, with one group for positive patients and another for negative.

3 Finding Number 1

For the models that were explored, the best results for each are displayed in Table 1. This shows that the more complex models performed better than the baseline set by Logistic Regression, as well as that achieving the 90% accuracy was a difficult challenge that only the Decision Tree model achieved. The accuracy results are presented based on the test data, as well as the cross-validated accuracy on the training data.

Model	Accuracy (%)	Cross-Validated (%)	Sensitivity (%)
Logistic Regression	85	X	X
Support Vector Machine	89	X	X
Decision Tree	90	93	88
Random Forest	89	X	X
K-Nearest Neighbour	89	X	X

Table 1: Summary of the different model performance

The other metric of interest in addition to accuracy is sensitivity. This measures how many children with conductive conditions are correctly identified as part of the positive class. The higher the sensitivity, the less positive patients that slip through the gaps. For the Decision Tree with the best accuracy, its sensitivity is 88%, which means it only classifies 12% of positive patients as negatives, or false negatives. Both the accuracy and the sensitivity are important metrics to consider for the final model.

In determining the final model, there are multiple criteria that the models were evaluated against. The first criteria is performance, as a higher accuracy leads to better outcomes for patients. This eliminated Logistic Regression as it has the lowest results. The remaining four models performed at a similar level, leading to the second criteria, explainability.

The client not only wanted to improve the accuracy of the model used, but wanted to better understand WBT data and how it relates to conductive conditions. The model needs to provide insights into why it made a certain prediction, or what data it is using (or not using) to do so. This criteria eliminated the Random Forest and K-Nearest Neighbour models as they are both black box models and their results are not easily interpretable.

The remaining model architectures, Support Vector Machines (SVM) and Decision Trees, perform well and are inherently explainable. The Decision Tree model was selected over the SVM as it has slightly better performance, its explainability is slight better due to the ability to extract feature importances directly from the tree, and it generates as simple graphic as seen in Figure 1 that allows clinicians identify why a certain outcome is being predicted. In addition, Decision Trees do not require scaling which results in a simpler processing step compared to the SVM model.

There are limitations to Decision Trees however, as they are unstable due to being only one tree (as compared to a Random Forest), and future data may cause changes in the features and splits used [2]. This may change the understanding of what is important in the data, and more data needs to be collected to ensure that the tree stability is established.

4 Looking Ahead

With a model selected that improves on the current results, the next step in this project is to collect more data. This will allow the model to learn more of the trends from a wider population and increase its stability of predictions on unseen data. More data can also lead to exploration of more advanced models such as Neural Networks that could learn complex relationships between WBT data and conductive conditions in children.

Using the current results, the implementation of the model in the data analysis process will allow for its improved accuracy to be realised. If this is automated as has been done for this project, clinicians will be able to save time doing the data analysis, and have more time to spend with their patients. The TyTy project is not aiming to replace audiologists in the process, but to become a tool for them to work smarter and use their amplify the use of their expert knowledge by identifying where in the data they should be focussing on.

5 Contribution

To Add: Gantt Charts

Member	Number	Project Tasks	Skills
Di Yao	22795234	1. Research Wideband Absorbance Data and scope the project 2. Review code and approve changes using Pull Requests 3. Perform modelling using K-Nearest Neighbour and then Random Forest	Python Machine Learning Git and GitHub
Karan Rebello	22868277	1. Research Wideband Absorbance Data and scope the project 2. Review code and approve changes using Pull Requests 3. Perform modelling using Random Forest	Python Machine Learning
Anitha Raghupathy	22773933	1. Research Wideband Absorbance Data and scope the project 2. Review code and approve changes using Pull Requests 3. Perform modelling using Support Vector Machines	Python Machine Learning
Cheng Nian	23053313	1. Data Cleaning and Transformations including pressure-matching function 2. Review code and approve changes using Pull Requests 3. Perform modelling using Support Vector Machines	Python Machine Learning
Aminul Islam	22884375	1. Data Cleaning and Transformations 2. Perform modelling using Logistic Regression	Python Machine Learning

Daniel Chegwidden	21282744	1. Create GitHub Repository and set up applicable automation and structure around Commits and Pull Requests 2. Manage code integration to ensure a workable code base was maintained and applicable functions and structures were in place to support consistent and reproduceable analysis 3. Review code and approve changes using Pull Requests 4. Perform modelling using Decision Tree	Team Leadership Meeting Organisation Python Machine Learning Git and GitHub Software Engineering

References

[1] W. H. Organization et al. *Global costs of unaddressed hearing loss and cost-effectiveness of interventions: a WHO report, 2017*. World Health Organization, 2017.

[2] Y.-Y. Song and L. Ying. Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, 27(2):130, 2015.