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

Contents

1 Aim

This project aims to apply machine learning techniques to wideband tympanometry (WBT) data in order to classify ears as either normal, or with conductive conditions such as Otitis Media. These techniques seek to replace the receiver operating characteristic (ROC) analysis that is currently performed, by increasing the performance and improving the interpretability of the results.

2 Background

The ear isn't just the hearing organ. It is a complex system of parts that not only allows humans to hear, but also makes it possible for humans to walk. The middle ear couples sound waves from the low impedance, air-filled outer ear to the high impedance, fluid-filled inner ear. The traditional method (226Hz tympanometry) for testing the function of the middle ear from 7 months old is through using a single tone at 226Hz to measure the acoustic changes in the ear. However, the traditional method has high rates of false positives (11%) and false negatives (25%) when detecting conductive conditions (Zielhuis et al. [4]). To tackle this clinical conundrum, an emerging technology (wideband tympanometry; WBT) has been used to generate data from 250Hz to 8000Hz, giving a more accurate picture of what is present in the middle ear, namely looking for conductive conditions.

The WBT technology measures the proportion of acoustic energy absorbed by the middle ear at the tympanometric peak pressure. Coupled with being measured at more than 60 different pressures, there is a large amount of data to analyse to understand what is happening in the middle ear. This methodology has been available for almost two decades [2] but has seen low uptake due to the limited ability of clinicians to interpret the result. Traditionally, the receiver operating characteristic (ROC) has been used to interpret WBT results, with the measure of success being the sensitivity and specificity metric.

This project seeks to apply machine learning to improve the effectiveness of the analysis by increasing specificity and targeting the interpretability of the results. By taking in the raw WBT data, processing can happen automatically, and the top-performing model can be used to predict the occurrence of conductive conditions. Coupled with simple visualisations and an easy-to-understand output, this would allow clinicians to diagnose different middle ear conditions with high accuracy and reduce unnecessary medical intervention such as surgery. There are approximately 2700 VTI surgeries conducted in WA public hospitals every year, at an estimated annual cost to the health service of $13.7 million. This project seeks to help reduce that number through reducing the false positive rate of conductive conditions in children.

3 Value Proposition

The use of machine learning in analysing the WBT data has the following key benefits:

1. Improving on the performance of existing methods. The current best area under the

ROC results are 77%, which is the target to achieve for the machine learning models. We will also use accuracy as a good classification measure, as well as sensitivity and specificity to fine-tune the models' performance.

2. Increasing the interpretability of the results. The current ROC analysis is often difficult for clinicians to interpret and apply its outputs to their patients. The machine learning models seek to output a clear and simple result to empower their application for the benefit of patients. There is a large amount of data collected, and using feature selection to identify valuable and unnecessary data points would give clinicians greater understanding of what is important.

3. Enable fast and reproduceable analysis. By creating a trained model, machine learning can quickly analyse WBT data from new patients and provide clinicians with an instantaneous classification of their patients' ears for conductive conditions. Coupled with an automated processing pipeline, raw data collected from patients can be directly input into a process that can generate the interpretable results that clinicians require.

4 Deliverables

The deliverables for Tytonidae Tympanometry are as follows:

1. A machine learning solution that takes in the WBT data, and any relevant demographic data, to classify a patient's ears as either normal or with conductive conditions. The output of this solution will not only be the classification, but also a visual analysis of how the prediction was made.

2. A code repository on GitHub which explores different machine learning methods to identify the best performing one, as well as processing steps, that can be used for further or improved analysis. This code repository and any outputs will use Python to generate the results.

3. A report that summarises the data, the processing steps that occurred, the models that were explored, and the results that were achieved by each.

5 Methods

For this project, Python has been selected as the programming language of choice, with Pandas and Scikit-Learn being the key Python libraries for data manipulation and machine learning. These were chosen for their versatility and simplicity, to support the ability to easily interpret the results.

Looking at previous work done with wideband absorbance data and machine learning by Grais et al. [1] and Sundgaard et al. [3], there appears to be a preference for more complicated black-box models such as deep learning and convolutional neural networks (CNN). These may provide better results, but they conflict with the requirement to be explainable to clinicians. With this in mind (and still aiming for the best performance), the following machine learning techniques have been selected for the initial analysis:

1. Logistic Regression - LR is a statistical technique that looks at the probability of a

record (a patient in this situation) belonging to either the positive class (with a conductive condition) or the negative class (without a conductive condition - a normal middle ear). As logistic regression is a simple model compared to the others, it provides a baseline result upon which to compare the other results with data that has gone through the same processing steps.

2. Support Vector Machine - SVM is a supervised learning technique that performs classification by creating a division between the different classes present in the data. The wider this division, and the less data that falls on the wrong side, the better the classification results. SVM is effective in high-dimensional space, which is useful for WBT data as it is collected across 16 frequencies, plus the additional demographic features that are added in.

3. Decision Tree - DT is one of the most explainable machine learning techniques, where the model is made up of branches and leaf nodes. Each branch represents a test, and each leaf node is the resulting class that is being predicted. This allows a logic flow to be followed down the tree, seeing which features are contributing to the prediction of the class. The features that have a greater effect of the prediction, the higher they are up in the tree, which will give insights to clinicians as to why the model is providing a certain result.

4. Random Forest - RF is a ensemble model that incorporates many decision trees and averages out the results of each individual tree. As this incorporates multiple trees, it inherently takes longer than a single tree, but a wisdom-of-the-crowd approach results in a decrease in variance of the results. Even though each tree is easily interpretable, the overall forest loses this due to the nature of averaging across a large number of trees.

All of these models will be compared across a range of metrics, and with the existing ROC results, to see which is the best performing model for incorporation into an automated process. The data that is being passed into these models will incorporate different levels of feature selection to identify those features which add predictive value to the model, as well as one of the following record selection methods:

1. All data - this will incorporate all of the WBT data across all pressures to see if there is any predictive power in any of the data that was collected.

2. Matching Pressures - only the records where the pressure recorded matches the tympanometric peak pressure which is recorded as the Adult Absorbance feature in the data.

Additional models such as K-Means Clustering or Neural Networks may be explored given the time constraints and the performance of the above four models in order to deliver the best results.

6 Project Management

This project will be completed over twelve weeks which is divided into three Phases (Processing, Modelling, and Reporting). The six team members will each contribute approximately 8 hours per week in Phase One and Phase Three, and 4 hours per week in Phase Two. This results in 480 hours spent over the 12 weeks. With an estimated cost of $40 per hour, this project is worth $19,200. The distribution of work amongst the team

and across the Phases can be found in the Gantt Charts in Appendix A.

References

[1] E. M. Grais, X. Wang, J. Wang, F. Zhao, W. Jiang, Y. Cai, L. Zhang, Q. Lin, and H. Yang. Analysing wideband absorbance immittance in normal and ears with otitis media with effusion using machine learning. *Scientific Reports*, 11(1):1–12, 2021.

[2] M. K. Park. Clinical applications of wideband tympanometry. *Korean Journal of Otorhinolaryngology-Head and Neck Surgery*, 60(8):375–380, 2017.

[3] J. V. Sundgaard, J. Harte, P. Bray, S. Laugesen, Y. Kamide, C. Tanaka, R. R. Paulsen, and A. N. Christensen. Deep metric learning for otitis media classification. *Medical Image Analysis*, 71:102034, 2021.

[4] G. Zielhuis, G. Rach, and P. Van Den Broek. Screening for otitis media with effusion in preschool children. *The Lancet*, 333(8633):311–314, 1989.

7 Appendix - Contributions

7.1 Di Yao

I have worked closely with the team on the wideband tympanometry project, and all my team members have been very accommodating for me to join meetings virtually and communicate with them over Teams. Our team leader, Daniel, has been delegating tasks on GitHub and each of us could track what we need to do and review each other's work. There is a weekly meeting scheduled and I am able to join virtually and contribute.

After our team discussions and meeting with the client, our team has developed a better understanding for the dataset and what we need to do. The aim of this project is to identify a machine learning tool that could automatically detect ears with conductive conditions and more importantly, improve accuracy from the current method of ROC curve.

We were divided into subgroups and my group worked on drafting the proposal. I have worked on the aim, background and value proposition. We had team meetings within our sub team and finalized the proposal draft for the other group to review. I now am clear with what exactly we need to deliver and what models we will be testing to find the best model. I have completed the machine learning unit, which I believe it will be very helpful in modelling the WBT data to improve accuracy in detecting abnormal ears. Same to other team members, I will work on one of the models we have decided and compare results with the ROC results and other models. As we have outlined in the proposal, we will ensure interpretability of our deliverables and machine learning model. This is what I want to focus in our final report, where I could explain the modelling process and the results in a non-technical way. I have enjoyed project and working with my team member so far, very looking froward to the next phase of our project.

7.2 Karan Rebello

Having finished most of the major units, I believe I am well-equipped to work on a project that gives me the chance to be better prepared for the industry. However, I realise that there is much to learn yet. I hope to utilise this opportunity to not only make the best use of my strengths but also identify my weaknesses and start working on them. One such example would be familiarising myself with GitHub since this project is more about collaboration than anything else. Familiarising myself with GitHub will be a huge learning curve as going forward irrespective of where I work, I realise that this will be the platform used to integrate code developed by fellow colleagues.

Coming to my strengths and contributions to the project, I would bring to the table my skills in python and ML as well as data analysis that would help me understand the data and accordingly be able to build models tailored to the client's needs which deliver optimum results and are easy to interpret. Irrespective of the "issues" raised on Git and who it is assigned to, I would like to work on multiple models and review their code (as they would mine) to be able to increase my comfort in ML and be able to bounce off ideas with my teammates about what else could be done to further fine tune our models. My communication and written skills will also come in handy while documenting our findings and processes and writing the final report. For now I have contributed towards the proposal and proof read it as well as reviewed the code written by my peers.

Lastly, but most importantly this project does not only test our skills but challenges us in terms of collaboration. All my teammates, both the entire group and the subgroup that I work in, seem very knowledgeable and have been very warm and welcoming. Irrespective of work, background, location, and personal commitments each of us have, we seem to adjust our routines and come together to work as a team helping each other out whenever and wherever required. I look forward to working with them and learning from them as much as possible as well as share as much as I can.

7.3 Anitha Raghupathy

In the past weeks, I've enjoyed working with my teammates collaborating through both virtual and face-to-face meetings for discussion and GitHub for code integration. It is exciting to work on this Capstone Project as it has given an opportunity to not just utilise my technical skills but also enhance soft skills like teamwork and effective communication.

In Phase 1 of this project, I was assigned to Team A along with Di and Karan by our team leader Daniel. Knowledge acquired through Agile Web Development unit has enabled me to contribute to the deliverables and Gantt Chart part of the proposal document. In the following Phases I seek to work on data modelling part to train WBT data. I have completed Machine Learning unit in previous semester, and I believe it will help me with this part of the project to generate models and perform fine tuning to optimise performance.

It's been a pleasure working with very co-operative and highly knowledgeable teammates, guided by our amazing supervisor Robyn. Looking forward to working on the next phase.

7.4 Cheng Nian

With data processing experience and machine learning skills, I am able to contribute to most parts of the project, under teamwork and collaboration with my group mates. I have done my part of data processing, which consists of two parts: 1. pressure matching, which is to find the exact pressure that gives the peak compliance in the tympanometry result, and then group them for further exploration; 2. manually check data merging: since parts of the two datasets are missing, when doing the inner join of datasets, there could be data loss. Thus I manually checked both datasets and made notes of data missing. Also as a data science student, I have no prior knowledge of tympanometry, so I did research to make sure I understanding the client requirements as well as how to proceed with my technical skills.

Our group chooses GitHub as the collaboration tool, and our team leader has created a Git repository for us to work collaboratively. Clarity is important. I am goal-oriented and so I follow the team agendas and guidance, deal with the tasks assigned and make sure delivering my work on time to make thing goes smoothly.

I am always keen to learn, and I believe the more effort I put in, the more I can gain from doing this project. With weekly meetings as a group and client meetings, I can also develop communication skills and time managing skills, which are vital during the collaboration work.

7.5 Aminul Islam

I am working in a team of six members with an objective to apply Machine Learning to predict hearing loss using Wideband Absorbance data. we have nominated a team leader who built a git repository for this team in such way that everyone can contribute to the project development. I think that good teamwork is the key to success in project activities when time and resources are limited. As everyone had their own point of view, many different ideas could be produced, and I found the energy of group participation made me feel more energetic about contributing something. It is an exiting working environment from where I will able to develop skill related software development. At the inception of the project development, I work with the features selection, data type selection of the Wideband Absorbance data at on Phase 1.

Each team member might have a different set of skills they can bring to the planed weekly meeting. I found very it very important to take part in meeting discussion and take relevant notes to my designated jobs. I will try to solve raised Issues on GitHub for tasks that need to be accomplished to allow the other members to easily identify their activities. I will use python and machine learning skill to the project advancement. Group work on a projects help me to work on my communication and conflict management skills. These are required in group work projects include speaking in turn, speaking up when I have ideas, actively listening to other team members' contributions, and crucially making compromises for the good of the team.

I strongly believe that I am a flexible in nature. I am interested in development of a effective machine learning tools with efficient coding which in turn can bring an breakthrough in medical diagnosis for Ear treatment.

7.6 Daniel Chegwidden

For this project I have taken on the role of team leader which works well with my software engineering skills, such as Git. As the leader, I set up the GitHub repository so that everyone could easily start working straight away, as well as writing guides for the team to follow for those that are less familiar with these technologies. I have walked the team in our developer meetings and individually through these so that the technology is not a barrier to anyone contributing.

I am fairly organised and so I create the agendas for each of our developer and client meetings, as well as take minutes so those that could not attend are able to keep up with what was discussed. I also liaise directly with the client to ensure that we understand the requirements and meet regularly to discuss progress. This involves some coordination with availabilities and balancing in-person and virtual team members.

I raise Issues on GitHub for tasks that need to be completed to allow the other members to easily identify what they are required to do. I am familiar with Python and so I review most work that is completed to ensure that it is compatible with the current working code, and then merge it into the main code that will form the basis of our solution. My Python skills also extend to machine learning and I have developed models using all of our proposed techniques. This allowed me to contribute to proposing these models, as well as being able to develop and review them, to ensure that they are working as intended.

My plan for this team is to develop their software engineering skills through the use of Git, and their Python development skills through building the solution. I am also a believer in collaborative development, so I split up the team in two to work on Phase 1 which is this proposal document and the data processing step. For Phase 2 and 3, these teams will switch so people get the chance to work with different people.

I am interested in deployment, and I am guiding the team towards writing high-quality, reproduceable code, to be translated into a working machine learning solution that the client can utilise as a clinician.

8 Appendix - Gantt Charts

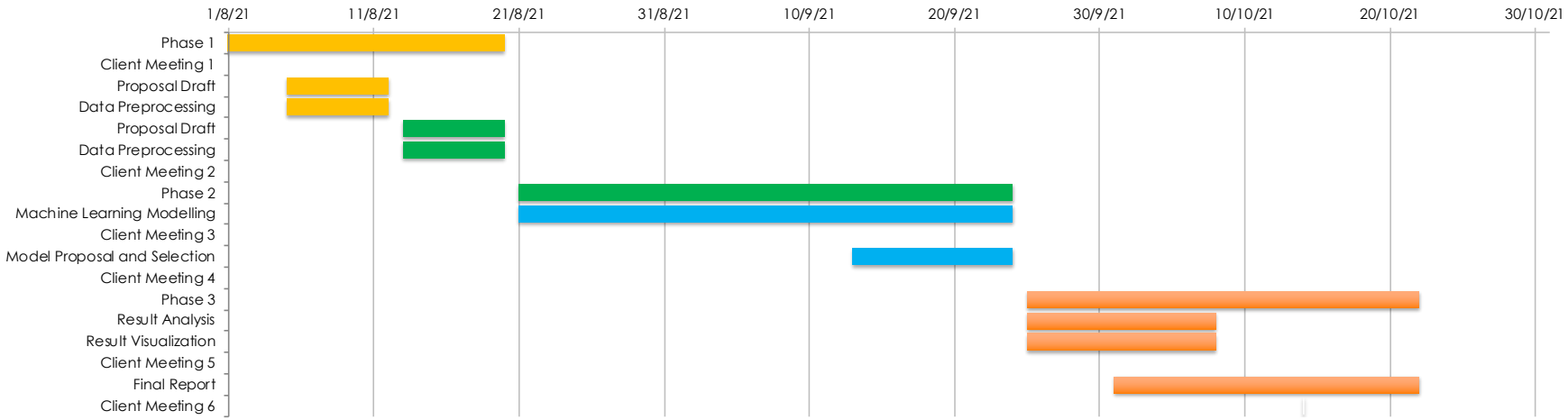
PROJECT NAME	PROJECT SUPERVISOR	START DATE	END DATE
Tytonidae Tympanometry	Robyn Choi	26-Jul	22-Oct

OVERALL PROGRESS
23%

PROJECT DELIVERABLE	1. ML Tool 2. Code 3. Final Report
TEAM MEMBERS	Daniel Chegwidden, Cheng Nian, Karan Rebello, Aminul Islam, Anitha Raghupathy, Di Yao

TASK NAME	RESPONSIBLE	START	FINISH	DURATION (DAYS)	STATUS	COMMENTS
Phase 1		26/7/21	20/8/21	25	In Progress	Team A - Karan, Di, Anitha Team B - Daniel, Cheng, Aminul Team Leader - Daniel
Client Meeting 1	Team Leader	28/7/21	28/7/21	0	Completed	
Proposal Draft	Team A	5/8/21	12/8/21	7	Completed	
Data Preprocessing	Team B	5/8/21	12/8/21	7	Completed	
Proposal Draft	Team B	13/8/21	20/8/21	7	Complete	
Data Preprocessing	Team A	13/8/21	20/8/21	7	In Progress	
Client Meeting 2	Team Leader	19/8/21	19/8/21	0	Complete	
Phase 2		**21/8/21**	**24/9/21**	**34**	Not Started	
Machine Learning Modelling	Entire Team	21/8/21	24/9/21	34	Not Started	
Client Meeting 3	Team Leader	2/9/21	2/9/21	0	Not Started	
Model Proposal and Selection	Entire Team	13/9/21	24/9/21	11	Not Started	
Client Meeting 4	Team Leader	16/9/21	16/9/21	0	Not Started	
Phase 3		**25/9/21**	**22/10/21**	**27**	Not Started	
Result Analysis	Team C	25/9/21	8/10/21	13	Not Started	
Result Visualization	Team D	25/9/21	8/10/21	13	Not Started	
Client Meeting 5	Team Leader	30/9/21	30/9/21	0	Not Started	
Final Report	Team E	1/10/21	22/10/21	21	Not Started	
Client Meeting 6	Team Leader	14/10/21	14/10/21	0	Not Started	

STATUS
Complete
Overdue
In Progress


1/8/21
11/8/21
21/8/21
31/8/21
10/9/21
20/9/21
30/9/21
10/10/21
20/10/21
30/10/21
Phase 1
Client Meeting 1
Proposal Draft
Data Preprocessing
Proposal Draft
Data Preprocessing
Client Meeting 2
Phase 2
Machine Learning Modelling
Client Meeting 3
Model Proposal and Selection
Client Meeting 4
Phase 3
Result Analysis
Result Visualization
Client Meeting 5
Final Report
Client Meeting 6

PROJECT NAME	PROJECT SUPERVISOR	START DATE	END DATE
Tytonidae Tympanometry	Robyn Choi	26-Jul	22-Oct

OVERALL PROGRESS
17%

PROJECT DELIVERABLE	1. ML Tool 2. Code 3. Final Report
TEAM MEMBERS	Daniel Chegwidden, Cheng Nian, Karan Rebello, Aminul Islam, Anitha Raghupathy, Di Yao

TASK NAME	RESPONSIBLE	START	FINISH	DURATION (DAYS)	STATUS	COMMENTS
Phase 1		**26/7/21**	**20/8/21**	**25**	In Progress	Team A - Karan, Di, Anitha Team B - Daniel, Cheng, Aminul Team Leader - Daniel
Developer Meeting 1	Team Leader	28/7/21	28/7/21	0	Completed	
Developer Meeting 2	Team Leader	2/8/21	2/8/21	0	Completed	
Proposal Draft	Team A	5/8/21	12/8/21	7	Completed	
Data Preprocessing	Team B	5/8/21	12/8/21	7	Completed	
Client Meeting 1	Team Leader	5/8/21	5/8/21	0	Complete	
Developer Meeting 3	Team Leader	9/8/21	9/8/21	0	Complete	
Proposal Draft	Team B	13/8/21	20/8/21	7	Complete	
Data Preprocessing	Team A	13/8/21	20/8/21	7	In Progress	
Developer Meeting 4	Team Leader	16/8/21	16/8/21	0	Complete	
Client Meeting 2	Team Leader	19/8/21	19/8/21	0	Complete	
Phase 2		**21/8/21**	**24/9/21**	**34**	Not Started	
Machine Learning Modelling	Entire Team	21/8/21	24/9/21	34	Not Started	
Developer Meeting 5	Team Leader	23/8/21	23/8/21	0	Not Started	
Developer Meeting 6	Team Leader	30/8/21	30/8/21	0	Not Started	
Client Meeting 3	Team Leader	2/9/21	2/9/21	0	Not Started	
Developer Meeting 7	Team Leader	6/9/21	6/9/21	0	Not Started	
Developer Meeting 8	Team Leader	13/9/21	13/9/21	0	Not Started	
Model Proposal and Selection	Entire Team	13/9/21	24/9/21	11	Not Started	
Client Meeting 4	Team Leader	16/9/21	16/9/21	0	Not Started	
Developer Meeting 9	Team Leader	20/9/21	20/9/21	0	Not Started	
Phase 3		**25/9/21**	**22/10/21**	**27**	Not Started	

STATUS
Complete
Overdue
In Progress

Result Analysis	Team C	25/9/21	8/10/21	13	Not Started	
Result Visualization	Team D	25/9/21	8/10/21	13	Not Started	
Developer Meeting 10	Team Leader	27/9/21	27/9/21	0	Not Started	
Client Meeting 5	Team Leader	16/9/21	16/9/21	0	Not Started	
Final Report	Team E	1/10/21	22/10/21	21	Not Started	
Developer Meeting 11	Team Leader	4/10/21	4/10/21	0	Not Started	
Client Meeting 4	Team Leader	7/10/21	7/10/21	0	Not Started	
Developer Meeting 12	Team Leader	11/10/21	11/10/21	0	Not Started	
Client Meeting 6	Team Leader	16/9/21	16/9/21	0	Not Started	
Developer Meeting 13	Team Leader	18/10/21	18/10/21	0	Not Started