Iowa Initiative for Artificial Intelligence
Final Report

Project title:	Leveraging Artificial Intelligence to Unravel the Radiogenomic Architecture of Hearing Traits
Principal Investigator:	Ishan Bhatt, Ph.D., Communication Sciences and Disorders
Prepared by (IIAI):	Avinash Reddy Mudireddy, Jacob Kitzmann, Valerie Ingalls
Other investigators:	
Date:	4/13/2026

Were specific aims fulfilled:	Yes
Readiness for extramural proposal?	
If yes … Planned submission date	
Funding agency	
Grant mechanism	
If no … Why not? What went wrong?	

Brief summary of accomplished results:

Auditory disorders such as age-related hearing loss (ARHL), difficulty discerning speech in background noise (speech in noise) and tinnitus affect over one third of adults over 60 [1]. With many advancements in treatment, the neurophysiological basis of these conditions remains unclear. We propose an artificial intelligence approach for prediction of these auditory disorders utilizing data from UK Biobank leveraging both genetic data and physiological variables of the brain kept as MRI metadata. We aim to create three predictive models, targeting ARHL, speech in noise, and tinnitus, respectively. Features will be selected from the top importance features of parallel radiologic data-based and genetic data-based pipelines. Shapley values will be used to determine feature importance, then features will be combined to create three comprehensive predictive networks leveraging both neurophysiological and genetic data. Final combined models successfully predicted above chance for ARHL (highest AUC-ROC = 0.69) and speech in noise (highest AUC-ROC = 0.60) but not tinnitus. Our results represent a promising beginning with clear avenues for future research.

Research report:

Aims (provided by PI):

Age-related hearing loss (ARHL), tinnitus (ringing in the ears or head), and speech-in-noise (SIN) deficits are common, complex, and heritable auditory conditions with no medical cure. ARHL is the leading cause of hearing loss, affecting over 1.4 billion adults worldwide, and this number is expected to rise significantly in the coming decades. Tinnitus and SIN deficits are frequent comorbidities of ARHL, further compounding the burden on those affected. Tinnitus alone impacts over 750 million people, with 120 million experiencing severe, debilitating symptoms that disrupt sleep, cognition, and mental health. About 26 million US adults report SIN deficits, often described as difficulty hearing in noisy environments despite having normal hearing. Collectively, these conditions increase the risk of cognitive decline, dementia, and insomnia, contributing to a growing economic burden due to ongoing healthcare needs after diagnosis.

Despite the high prevalence of these auditory conditions, their biological underpinnings remain poorly understood, hindering the development of effective treatments. While genetic studies have begun to uncover their complex architecture, standard statistical approaches primarily focus on linear, additive effects, overlooking the intricate interactions between genetic and radiological factors that likely drive these conditions. As a result, the influence of genetic variants on brain structure and function contributing to ARHL, tinnitus, and SIN deficits remains largely unexplored.

Recent advances in artificial intelligence (AI)—including generative pre-trained transformer (GPT) models and explainable AI (XAI)—hold significant promise for modeling complex traits. Unlike traditional methods, GPT models can capture linear, nonlinear, and complex associations, offering a more comprehensive approach to risk prediction. XAI techniques can then unravel these models, providing insights into the relationships between genetic markers, brain structure, and disease risk. However, robust GPT models that specifically predict ARHL, tinnitus, and SIN deficits by integrating genetic and radiological data are still lacking. There is a critical need to create AI-powered radiogenomic models to better understand and predict these auditory conditions.

Our long-term goal is to develop clinically useful AI-based tools that enable personalized medicine and support the discovery of novel therapies for ARHL, tinnitus, and SIN deficits. The primary objective of this proposal is to build and validate GPT models that integrate genetic and radiological risk factors across diverse populations to predict the risk of ARHL, tinnitus, and SIN deficits. Our specific aims are as follows:

Aim 1: To develop and validate a GPT predictive of ARHL with genetic and radiological profiles.

Aim 2: To develop and validate a GPT predictive of tinnitus with genetic and radiological profiles.

Aim 3: To develop and validate a GPT predictive of SIN deficits with genetic and radiological profiles.

<u>Final aims decided after project started:</u>

The overarching goal of this project is to leverage advanced AI techniques to uncover and model the neurophysiological basis of common auditory disorders through the development of predictive classification models. Specifically, we aim to develop systems for:

1. *Prediction of age-related hearing loss (ARHL)*
2. *Prediction of difficulty discerning speech in background noise (speech in noise)*
3. Prediction of tinnitus

Data:

We have access to genetic datasets from the UK Biobank (N=500,000), Haplotype Reference Consortium (N=32,488), Trans-Omics for Precision Medicine (N=138,134), Human Genome Diversity Project (N=929), and NIH All of Us (N=254,700). We will use radiological data from 100,000 UK Biobank participants.

Data was acquired from UK Biobank (UKB), a population-based database of health information containing over 500,000 citizens of the United Kingdom aged 40–69. Genomic data for over 90 million

single nucleotide polymorphisms (SNPs) is available for all individuals in the database, in addition to various health survey data. The outcome measures for this study—ARHL, speech in noise, and tinnitus—were measured via self-report. A subset of the cohort also received MRI scans. Several modes of MRI were utilized, including resting functional MRI (fMRI), T1-weighted structural MRI, diffusion MRI, and T2 FLAIR MRI. All participants used in the training and testing of the final models in the study had both genetic and radiologic data available.

Radiologic data consisted of the metadata of each subject's acquired DICOM series. Metadata included many physiological calculations from the scan, including volumes of brain segments, blood flow, and diffusion metrics. Many of these features indicate both anatomical and physiological differences in patients, with features such as fractional anisotropy—a measure of anisotropic water flow in white matter—and mean orientation having demonstrated direct relationships to cognitive function.

The UKB's genotype data is derived from approximately 850,000 directly sequenced variants. These SNPs are passed through an imputation pipeline, resulting in the final set of over 90 million SNPs. For this study, a subset of SNPs associated with age-related hearing loss in Trpchevska et al.'s genome-wide association study meta-analysis (meta-GWAS) were selected for analysis [2]. Originally, we planned to use approximately 650,000 SNPs achieving $p < 0.05$ significance in the meta-GWAS. However, attempting to use that many SNPs for the over 400,000 individuals included in the genomics-only branch of our study created difficulties with computer memory. One TB of memory was insufficient to load the full dataset, even after performing PCA for feature reduction, thus leading us to narrow the scope of the study. We raised the p-value threshold for inclusion to 5×10^{-6}, the traditional threshold for suggestive significance in GWAS. This resulted in a final set of 7,358 SNPs that were used in this study.

SNP effects are difficult to interpret, as they are often small and biologically nebulous. Therefore, to improve the interpretability of our results and to control for multicollinearity, i.e., linkage disequilibrium, among our genetic factors, we performed gene annotation using ANNOVAR to map SNPs to their associated genes. We then performed principal component analysis (PCA) on the SNPs for each gene to derive a set of gene-specific principal components (GPCs) for each gene. Because genes have comparatively known functional roles in the body, this grouping and PCA increased the interpretability of our results. SNPs with more than 1% missing data were excluded prior to performing PCA. We used a mean imputation pipeline to replace any remaining missing values prior to PCA. GPCs explaining more than 99% of the variation for each gene were selected for final inclusion. This resulted in a final set of 881 GPCs used in this study.

AI/ML Approach:

Feature selection for the final network was executed using parallel development of radiologic-only and genetic-only classifiers. Each classifier was created through a pipeline of machine learning models progressing from simpler to more complex. Models included:

- Logistic Regression
- Random Forests Classification
- Extreme Gradient Boosting (XGBoost)
- Multilayer Perceptron

- Residual Network (ResNet)
- Capsule Network (CapsNet)

Shapley values of each feature were acquired from the parallel pipeline's highest-performing models based on area below the receiver operating characteristic curve (ROC-AUC). The Shapley values allowed the team to select the most influential features from each pipeline. The top-importance features were used to create comprehensive radiologic and genetic networks to predict each aim.

Experimental methods, validation approach:

Beginning data preprocessing, the dataset consisted of 141,992 subjects (rows) with 2,683 features (columns). Considering only subjects with completed MRI trials filtered the dataset to 31,453 subjects. Only the 100 most important features from the genetic and radiologic pipelines were used, reducing the feature size to 200.

Some subjects had varying MRI trials completed, which led to missing data in the dataset. A threshold of 20 percent missing information in a column was determined to be optimal for balancing data integrity and dataset size. Any column missing more than 20 percent of its data was removed. Remaining missing values were imputed with the mean of their respective column. This did not affect the combined radiologic and genetic pipelines, which utilized only the 100 most significant values from each. However, the individual pipelines were affected, and many columns were removed.

Some data was too highly ranged for a standard scaler to be effective without introducing variable importance. This was avoided with a power transformation. Any features that had a range above 10,000 were power transformed to allow for evenly weighted scaling to finish postprocessing.

After all preprocessing steps, a 56%-14%-30% train-validation-test split was applied to the data, finalizing preprocessing with 2,049 training cases, 513 validation cases, and 1,098 testing cases. All splits contained 100 radiologic and 100 genetic features.

GridSearchCV from Scikit-Learn was employed for hyperparameter optimization in the logistic regression, random forest, and XGBoost classifiers. TensorFlow's implementation of logistic regression was used with L2 regularization and an inverse regularization strength (C) of 0.05. A random forest classifier, also from TensorFlow, was configured with 100 estimators and a balanced subsample class weight. XGBoost's XGBClassifier was used with a binary logistic objective function. Neural networks were optimized through manual tuning and trial and error.

A multilayer perceptron (MLP) consisting of three hidden layers with 512, 128, and 64 units, respectively, was implemented. Batch normalization was applied to each layer, with ReLU used as the activation function, except for the output layer, which employed a sigmoid activation. The MLP was trained using the Adam optimizer with binary crossentropy loss.

A residual network was implemented using TensorFlow, comprising an initial dense layer of 128 units followed by two residual blocks, each with 128 units. Batch normalization and a dropout rate of 0.3 were applied, and ReLU activation was used throughout, with a sigmoid activation function at the output. The Adam optimizer and binary crossentropy loss were used for training.

Finally, a capsule network was implemented using a custom architecture. The model began with a dense layer of 64 units with L2 regularization, followed by batch normalization and ReLU activation. Dropout was applied before reshaping the output into 8 capsules of 8 dimensions each. A custom squash function was used to normalize capsule outputs, and the vector lengths were computed to serve as a proxy for class

probability. Batch normalization and dropout were applied to the capsule lengths, followed by a sigmoid-activated dense output layer. The model was trained using binary crossentropy loss and the Adam optimizer with gradient clipping (clipnorm = 1.0) to improve training stability.

The following Python packages were used for model implementation and evaluation: NumPy and Pandas for data loading and manipulation, Matplotlib and Seaborn for visualization, TensorFlow and Keras for neural network training and evaluation, Scikit-Learn for data preprocessing and hyperparameter optimization, and XGBoost for its extreme gradient boost classifier implementation.

Python 3.11.12 was used to develop the pipeline through The University of Iowa's Interactive Data Analytics Service (IDAS), a cloud-based high-performance computing cluster accessible through a Jupyter Notebook interface. The most powerful machine IDAS could offer was utilized. Our team used 32 CPU cores with 512 GB of memory and an NVIDIA L40S graphics card offering 48 GB of memory.

Results:

Before assessing model performances, a correlation matrix was created to assess feature similarity and correlation. The correlation matrix can be seen in Figure 1. There is a high level of correlation within the radiologic features, which are numbered beginning with "2." This is expected due to the anatomical correlation underlying the features. Neural pathways and brain volumes are highly correlated, and this must be taken into consideration when training models that assume independent data, such as neural networks. GPC features showed little to no correlation, which is expected from the dimensionality-reduction process of PCA, which is intended to create uncorrelated factors.

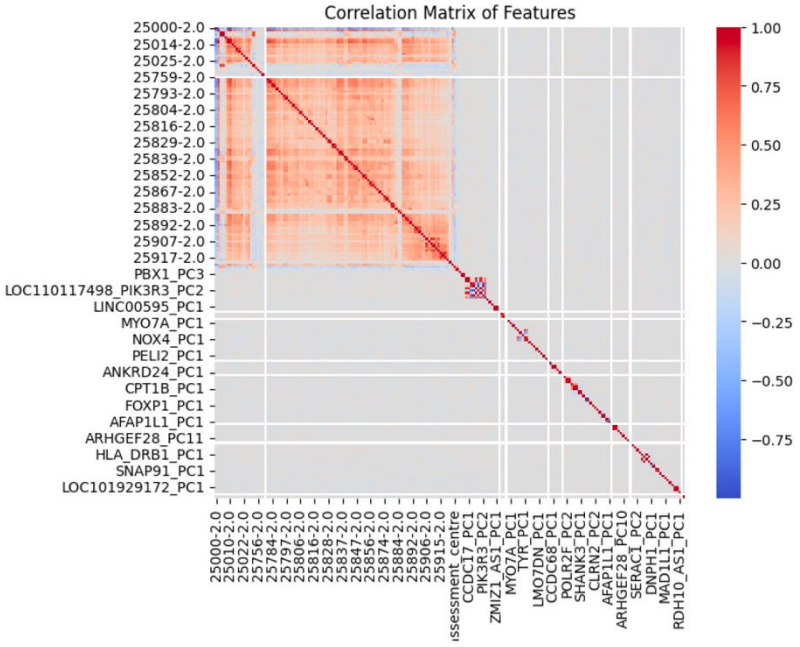

Figure 1. Correlation matrix of radiologic and genetic PC features

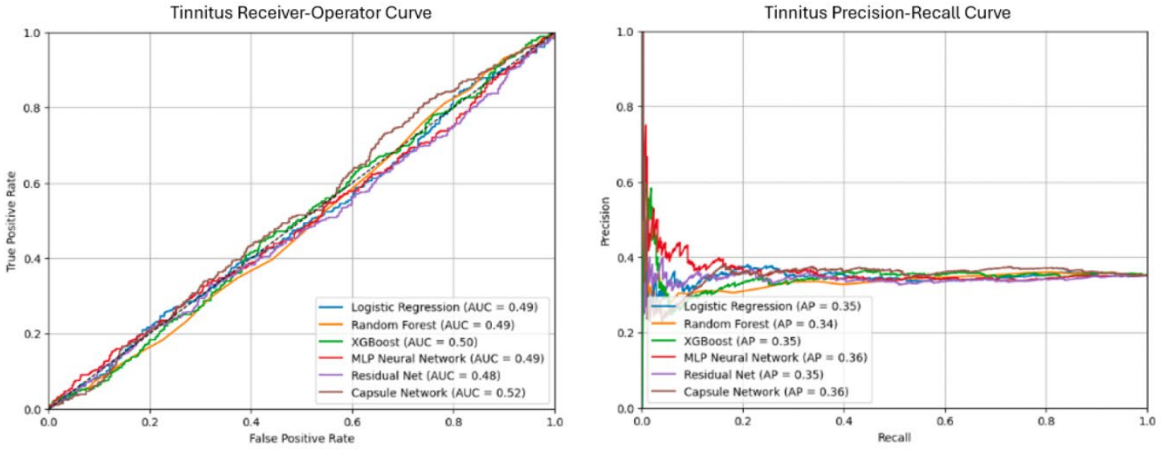

Figure 2. Receiver-operator curve and precision-recall curve for tinnitus prediction.

Combined genetic principal components and radiologic features showed no predictive utility for tinnitus. Individual radiologic and genetic pipelines did not show any increased predictive power compared with the combined pipeline. Both pipelines predicted binary random chance, with AUC approximately 0.5. All models also showed no predictive power from their precision-recall curves. There was some variation at low levels of recall; however, the precision flattened out to the percentage of true positives, indicating that the model was randomly guessing.

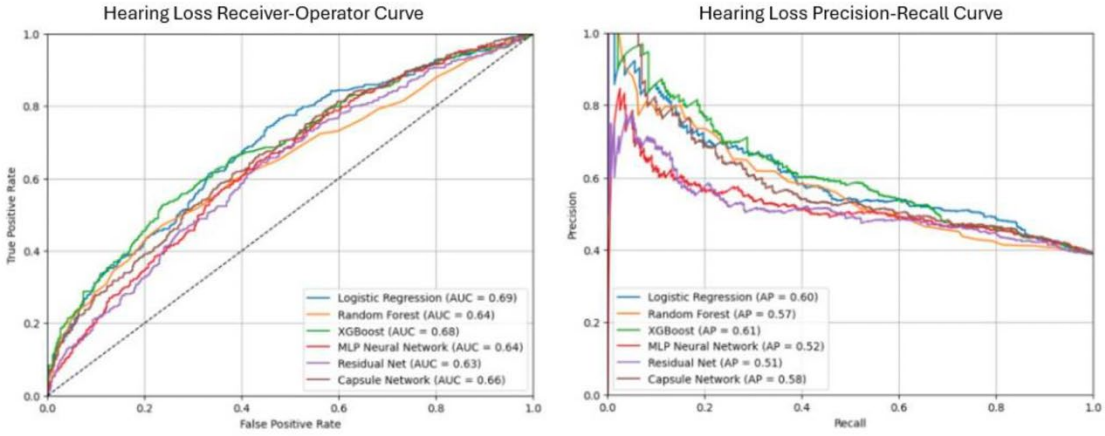

Figure 3. Receiver-operator curve and precision-recall curve for age-related hearing loss prediction

Combined genetic principal components and radiologic features yielded moderate predictive performance for hearing loss. The maximum AUC-ROC of the combined pipeline was 0.69 using logistic regression, slightly below the radiologic-only peak of 0.71 using CapsNet, but surpassing the genetic-only pipeline's AUC-ROC of 0.64 using logistic regression. Integration of genetic data also enhanced precision-recall performance, increasing the maximum AUC-PR from 0.58 in the radiologic-only CapsNet model to 0.61 in the combined XGBoost model.

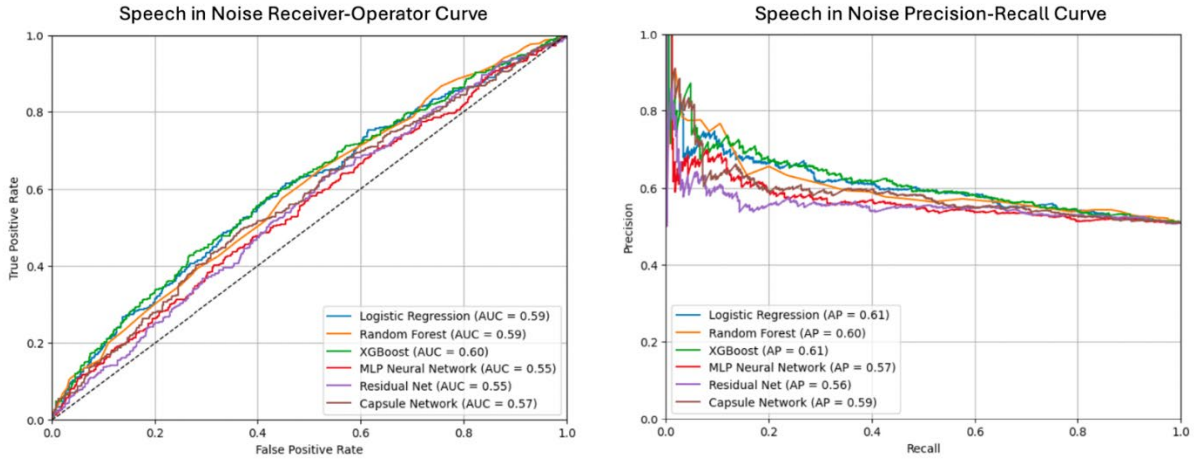

Figure 4. Receiver-operator curve and precision-recall curve for speech in noise prediction

Combined genetic principal components and radiologic features showed limited predictive power for speech in noise. AUC-ROC did not surpass 0.60 using XGBoost. This was a decrease in performance from the radiologic-only pipeline, which reached a maximum AUC-ROC of 0.61 using CapsNet, compared with the combined pipeline's maximum of 0.60 using XGBoost. However, integration of genetic data increased the pipeline's performance from a maximum AUC-PR of 0.58 in the radiologic-only CapsNet model to 0.61 in the combined XGBoost model.

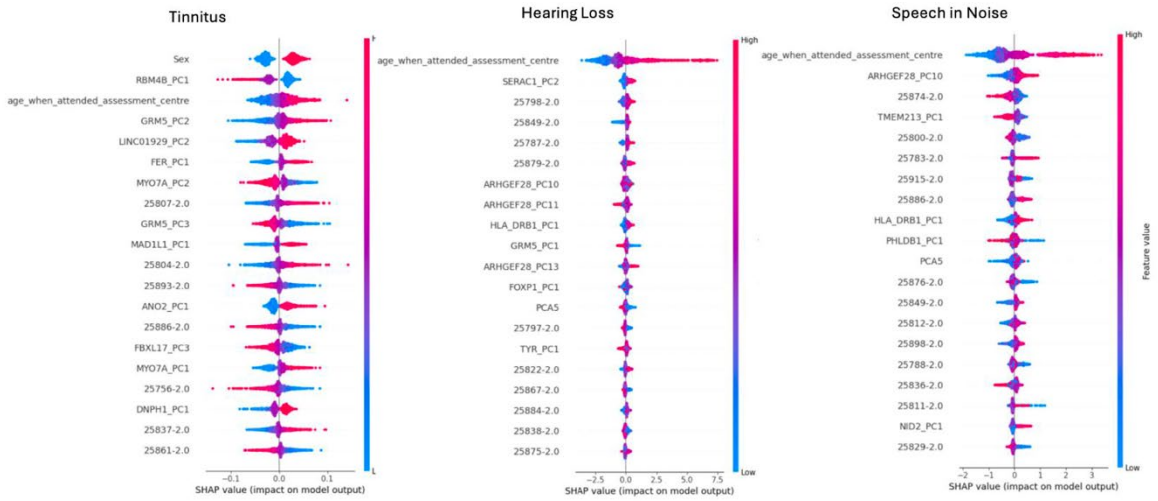

Figure 5. Shapley values of feature importance for each aim

Figure 5 shows the top Shapley values for each aim. Shapley values were calculated for the top-performing model of each aim: CapsNet, XGBoost, and XGBoost for tinnitus, age-related hearing loss (ARHL), and speech in noise, respectively. The few demographic features, sex and age, demonstrated high predictive power. Age showed very high feature importance in ARHL and speech in noise, with radiologic and genetic features supporting it. Features were much more balanced in tinnitus, which is expected because the model had little to no predictive power.

Discussion:

The results for this study demonstrate a positive, albeit limited, beginning to the work of radiogenomic modeling of auditory phenotypes. We were able to predict both age-related hearing loss (ARHL) and speech in noise phenotypes at a level above chance, and PR metrics indicate that the models were truly predicting, i.e., not defaulting to only case or only control predictions. Shapley values indicated that both genomic and radiologic features contributed to the combined models. Additionally, both radiologic-only and combined models outperformed the genomic-only models created using the narrowed subset of genomic features, and most combined models showed the best overall performance. We expect to see further improvement in our models with a more robust collection of genomic features.

Top MRI features revealed by Shapley values include gray-matter volumes in the frontal orbital cortex, right and left; parahippocampal gyrus anterior division, right; superior frontal gyrus, right; paracingulate gyrus, right; superior temporal gyrus; middle temporal gyrus, posterior division, left and temporo-occipital part, right; hippocampus, left; and cerebellar lobules I–IV, overall and left. Many of these regions have well-characterized roles in sensory processing, hearing, and memory.

Top genomic features revealed by Shapley values include GPCs related to SERAC1, ARHGEF28, GRM5, FOXP1, TMEM213, HLA_DRB1, TYR, and NID2. Many of these genes have previously been associated with hearing loss. SERAC1 is implicated in MEGDEL syndrome, which includes deafness as one of its main symptoms [4]. ARHGEF28, GRM5, TYR, and NID2 have all been implicated in previous GWAS of ARHL [e.g., 2]. Genes in the HLA region are associated with immune response and are known to be highly pleiotropic, making them commonly appear in genomic studies while remaining difficult to interpret. However, other genes, such as TMEM213, have not been seen in previous studies. The

importance of known hearing-related genes validates our approach, while the presence of novel genes is promising for generating novel insights.

Despite this, several important limitations present themselves in our results. First, we were unable to effectively predict tinnitus with either the individual or combined models. This may speak to the diversity of the tinnitus phenotype; if it has multiple anatomical generators that vary across individuals, then modeling using radiologic features may not be viable.

Furthermore, across all pipelines, logistic regression was one of the best-performing models. This suggests that features tended to have highly linear effects, with little interaction between them. If that is truly the case, then machine learning may not be the most appropriate approach for modeling these phenotypes. However, the selection of only the top genetic variations from the meta-GWAS, which was calculated using a linear approach, likely contributed to the linearity seen in our models. Some Shapley values, such as those for ARHGEF28_PC10 in the ARHL combined model, show that some small nonlinearity was present. It is very likely that the amount of nonlinearity would increase substantially with a wider array of genomic features. Therefore, we remain optimistic about the long-term viability of the approach used in this study, but further work is needed to determine whether it truly bears fruit.

The biggest pain point through the semester was handling the genomic data. Working with data at that scale was incredibly time-intensive due to both the computational burden and, at times, waiting for the Argon queue. We eventually had to narrow the scope to have a suitably finished product ready by the end of the semester. Ideally, this problem could have been foreseen and solved in advance. While there are merits to starting with straightforward approaches, the process would have been much simpler and smoother had we settled on a sufficiently robust data-handling method from the start.

Overall communication across the team was good. Both the domain PI and our AI supervisor were very helpful in answering questions and providing background information to understand the project.

Ideas/aims for future extramural project:

Preliminary work with a larger subset of genomic features improved performance in the genomic-only pipeline, suggesting that solving the memory issue and using the full set of genomic features should be a fruitful avenue for future work. This is the logical first step, as it falls within the original conception of the project. Selecting a new set of genomic features should hopefully improve the combined models.

Combined models were mostly created using the same hyperparameters and architectures as our genomic and radiologic individual models. In the future, tuning those models individually might be another avenue for improving performance.

The SNPs included in this study were selected from a meta-GWAS for ARHL. We know that different genes have been implicated in GWAS for speech in noise [5] and tinnitus [6]. Selecting genomic features from same-phenotype GWAS could improve model performance. Currently, the summary statistics for the tinnitus meta-GWAS are not publicly available, limiting our ability to take that approach for tinnitus.

We believe that implementing the full set of genomic features will result in substantial improvements to the genomic models. For those improvements to realistically propagate to the combined models, however, we will need to include a larger selection of features in those models. Having better features will of course help, but the expected gain in predictive utility for individual GPC features from increasing the number of SNPs used is minimal. Therefore, capturing the improvement from the more complicated genomic models is likely to require more complicated combined models. With these future combined

models, we will also be able to use Shapley values to examine feature interactions and determine whether there are gene-gene interactions or interactions between genomic and radiologic features.

References

[1] National Institute on Aging. Hearing Loss: A Common Problem for Older Adults. https://www.nia.nih.gov/health/hearing-and-hearing-loss/hearing-loss-common-problem-older-adults. Updated January 19, 2023. Accessed May 4, 2025.

[2] Trpchevska, N., Freidin, M. B., Broer, L., Oosterloo, B. C., Yao, S., Zhou, Y., Vona, B., Bishop, C., Bizaki-Vallaskangas, A., Canlon, B., Castellana, F., Chasman, D. I., Cherny, S., Christensen, K., Concas, M. P., Correa, A., Elkon, R., Metspalu, A., Nelis, M., … Nagtegaal, A. P. (2022). Genome-wide association meta-analysis identifies 48 risk variants and highlights the role of the stria vascularis in hearing loss. The American Journal of Human Genetics, 109(6), 1077–1091. https://doi.org/10.1016/j.ajhg.2022.04.010

[3] Sudlow, C., Gallacher, J., Allen, N., et al. (2015). UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Medicine, 12(3), e1001779. Published March 31, 2015. https://doi.org/10.1371/journal.pmed.1001779

[4] Finsterer, J., Scorza, F. A., Fiorini, A. C., & Scorza, C. A. (2020). MEGDEL Syndrome. Pediatric Neurology, 110, 25–29. https://doi.org/10.1016/j.pediatrneurol.2020.03.009

[5] Bhatt, I. S., Garay, J. A. R., Bhagavan, S. G., Ingalls, V., Dias, R., & Torkamani, A. (2024). A genome-wide association study reveals a polygenic architecture of speech-in-noise deficits in individuals with self-reported normal hearing. Scientific Reports, 14(1), 13089. https://doi.org/10.1038/s41598-024-63972-2

[6] Clifford, R. E., Maihofer, A. X., Chatzinakos, C., Coleman, J. R. I., Daskalakis, N. P., Gasperi, M., Hogan, K., Mikita, E. A., Stein, M. B., Tcheandjieu, C., Telese, F., Zuo, Y., Ryan, A. F., & Nievergelt, C. M. (2024). Genetic architecture distinguishes tinnitus from hearing loss. Nature Communications, 15(1), 614. https://doi.org/10.1038/s41467-024-44842-x