scientific reports

OPEN

Machine learning analysis of cardiovascular risk factors and their associations with hearing loss

Ali Nabavi[1,2], Farimah Safari[2✉], Ali Faramarzi[1,2], Mohammad Kashkooli[2], Meskerem Aleka Kebede[3], Tesfamariam Aklilu[4] & Leo Anthony Celi[5,6]

Hearing loss poses immense burden worldwide and early detection is crucial. The accurate models identify high-risk groups, enabling timely intervention to improve quality of life. The subtle changes in hearing often go unnoticed, presenting a challenge for early hearing loss detection. While machine learning shows promise, prior studies have not leveraged cardiovascular risk factors known to impact hearing. As hearing outcomes remain challenging to characterize associations, we evaluated a new approach to predict current hearing outcomes through machine learning models using cardiovascular risk factors. The National Health and Nutrition Examination Survey (NHANES) 2012–2018 data comprising audiometric tests and cardiovascular risk factors was utilized. Machine learning algorithms were trained to classify hearing impairment thresholds and predict pure tone average values. Key results showed light gradient boosted machine performing best in classifying mild or greater impairment (> 25 dB HL) with 80.1% accuracy. It also classified > 16 dB HL and > 40 dB HL thresholds, with accuracies exceeding 77% and 86% respectively. The study also found that CatBoost and Gradient Boosting performed well in classifying hearing loss thresholds, with test set accuracies around 0.79 and F1-scores around 0.79–0.80. A multi-layer neural network emerged as the top predictor of pure tone averages, achieving a mean absolute error of just 3.05 dB. Feature analysis identified age, gender, blood pressure and waist circumference as key associated factors. Findings offer a promising direction for a clinically applicable tool, personalized prevention strategies, and calls for prospective validation.

Keywords Cardiovascular risk factors, Hearing outcome, Machine learning

The latest statistics reveal that 72.88 million individuals in the United States experienced some form of hearing loss in 2019, representing a significant 22.2% of the total US population[1]. The most recent epidemiological data from 2024 reveals that approximately 1.5 billion people, which is almost 20% of the global population, are currently experiencing some level of hearing loss[2]. It is estimated that in 2050, approximately 698.4 million people will have moderate-to-severe hearing loss worldwide[3]. Hearing loss restricts social engagement and interactions while also potentially affecting learning, employment, mental health, and quality of life[4]. The increasing number of hearing loss cases underscores the urgency of early detection and prevention to minimize progress and enhance quality of life while reducing incidence[5].

Systemic diseases, such as cardiovascular disease (CVD), have been identified as risk factors for hearing loss. Unfortunately, once hearing loss has occurred due to these factors, it is irreversible[6,7]. Therefore, it is crucial to identify modifiable risk factors that can be addressed to reduce the risk of hearing loss. Recent reports suggest an independent strong association between hypertension[8], obesity[9], diabetes[10], and smoking[11] with hearing loss[7]. Also, studies have established a strong correlation between aggregate CVD risk factors and hearing loss[12,13]. Individuals with two or more CVD risk factors have an estimated 90% increased risk of experiencing hearing loss[7,14]. The general CVD risk profile score has also been associated with greater hearing loss in longer follow-up periods for both sexes[14]. The association between CVD risk factors and hearing loss has led to the development of several hypotheses about the mechanism behind it. Dysfunction in the blood supply of striae vascularis of the cochlea is believed to be the primary mechanism through which CVD risk factors impact hearing loss.

[1]Otolaryngology Research Center, Department of Otolaryngology, Shiraz University of Medical Sciences, Shiraz, Iran. [2]Student Research Committee, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran. [3]Global Surgery Policy Unit, Department of Health Policy, LSE Health, London, UK. [4]School of Medicine, Addis Ababa University, Addis Ababa, Ethiopia. [5]Laboratory for Computational Physiology, MIT Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [6]Information Systems, Beth Israel Deaconess Medical Center, Boston, MA 02215, USA. ✉email: farimasafari40@gmail.com

Sudden sensorineural hearing loss (SSHL) patients showed increased subclinical atherosclerosis as evidenced by greater carotid intima-media thickness in a recent systematic review of 61,000 patients, supporting the proposed vascular etiology underlying the development and progression of SSHL[15]. Mechanisms like strial atrophy, loss of spiral ganglion neurons, endothelial dysfunction, oxidative stress, and vascular inflammation, which affect the inner ear's vascular system, are other suggested etiologies[16,17].

The application of machine learning (ML) techniques offers significant potential for the fields of audiology and hearing science[18,19]. The use of computerized algorithms allows for the advancement of current methods of analysis by enabling the incorporation of multiple interconnected factors, detecting complex connections between clinical characteristics and results, and establishing highly predictive models using streamlined computational processes[20]. A crucial aim in hearing outcome research and early detection of hearing loss is to predict current hearing outcomes with a reliable model from a set of given associated factors, particularly CVD risk factors owing to their well-established relationship. Data-driven models used in this study permit algorithms to uncover patterns in the objective measurement and demographic data freely. This enables the construction of highly predictive models like neural networks without relying on potentially limited prior theoretical knowledge[21].

To date, no research has utilized ML techniques to predict current hearing outcomes based on main hearing loss risk factors as well as CVD risk factors. In this study, we use the National Health and Nutrition Exam Survey (NHANES) 2012–2018 to train and test the potential of different ML models. While comprehensive audiological evaluation encompasses multiple diagnostic procedures beyond pure-tone audiometry, early screening tools could help optimize the referral process in hearing healthcare pathways. Our objective is to develop a precise prediction model that can facilitate clinical decision-making based on the most accurate algorithm.

Method
Study design and data source

We conducted a retrospective analysis of data from the NHANES from 2012–2018. NHANES is an ongoing cross-sectional survey conducted by the National Center for Health Statistics to assess the health and nutritional status of adults and children in the US. The survey combines interviews, physical examinations, and laboratory tests using a complex, multistage probability sampling design to obtain nationally representative samples. The audiometry component measures hearing thresholds for speech and pure tone frequencies, allowing assessment of hearing function. NHANES data are publicly available at https://www.cdc.gov/nchs/nhanes/about_nhanes.htm.

From 28,874 NHANES participants, 20,988 were excluded due to age < 20 years, incomplete audiometric data, or self-reported hearing loss/related conditions. Exclusions covered hereditary or genetic factors, certain infections, specific medical conditions, medications, and noise exposure, including conditions such as congenital hearing loss, auditory neuropathy spectrum disorder, Meniere's disease, otosclerosis, and noise-induced hearing loss. The final sample comprised 7,996 participants ≥ 20 years with complete audiometric data and no self-reported hearing loss conditions. By excluding individuals with self-reported hearing loss conditions, we aimed to focus the analysis on predicting the onset of hearing impairment, rather than examining factors associated with pre-existing diagnosed hearing loss. Individuals with self-reported hearing loss may have different etiologies and risk factor profiles, which could introduce confounding factors and bias the predictive models. This sample was obtained by first removing those with missing audiometric data (17,849 excluded), then excluding those < 20 years (3,029 excluded), and after that excluding self-reported hearing loss/related conditions (110 excluded) (Fig. 1).

To develop our predictive model for hearing impairment, we conducted an extensive literature review to identify key demographic, clinical, behavioural, and environmental risk factors associated with CVD that
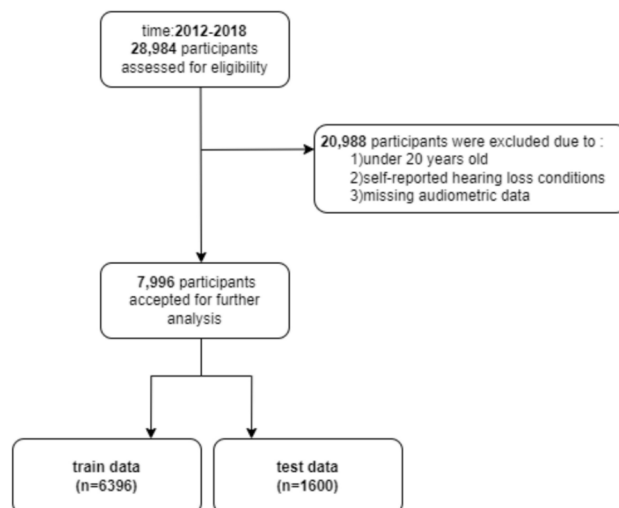

Fig. 1. Flow diagram of the cohort study.

may impact hearing health. The CVD risk factors, CVD health metrics, and CVDs were selected based on the guidelines developed by the American College of Cardiology (ACC), the American Heart Association (AHA), and the Framingham Heart Study[22,23]. Based on this comprehensive review, we selected 50 relevant variables consistently measured in the NHANES 2012–2018 dataset, allowing assessment of main hearing loss risk factors as well as CVD risk factors in a robust, nationally representative population sample (Table S1).

The demographic attributes included age, gender, race/ethnicity, poverty income ratio and education level. CVD risk factors and history captured self-reported diagnoses of conditions like diabetes, prediabetes, hyperlipidemia, and hypertension as well as related prescription medication use. Clinical biomarkers measured included glycosylated hemoglobin (HbA1c), fasting plasma glucose, serum lipids (HDL, LDL, total cholesterol), blood pressure, anthropometric data (body mass index, waist circumference), and cotinine levels. Dietary intake variables from 24-h recalls focused on components relevant to the Dietary Approaches to Stop Hypertension (DASH) diet. Physical activity domains included transportation, leisure activities, and exercise duration/intensity. Questions on smoking and alcohol consumption assessed lifetime use and current intake. Access to healthcare was represented by insurance status and provider counseling on conditions like hyperlipidemia and hypertension. Cardiovascular-associated factor data were obtained from the closest available clinical encounter in the NHANES dataset, which may not correspond to the exact day of the audiological evaluation, potentially introducing variability due to day-to-day fluctuations in these measures.

Primary outcome

We defined hearing impairment based on the pure-tone average (PTA) audiometry, calculated as the mean hearing threshold in decibels hearing level (dB HL) at 500, 1000, 2000, and 4000 Hz for better ear on NHANES audiometry testing. Aligning with American Speech-Language-Hearing Association (ASHA) guidelines, we categorized hearing thresholds into three distinct groups for analysis:

Normal vs. Abnormal Hearing: Threshold set at > 16 dB HL to identify deviations from normal hearing.

Slight vs. More than Slight Impairment: Threshold set at > 25 dB HL to distinguish slight/normal impairment from more significant impairment.

Severe vs. Less than Severe Impairment: Threshold set at > 40 dB HL to identify severe hearing loss[2,24]. Our primary analysis focused on the > 25 dB HL threshold to classify binary hearing impairment. Additionally, supplementary analyses applied the > 16 dB HL and > 40 dB HL thresholds to explore slight impairment and severe hearing loss, respectively.

Additionally, we developed ML regression models to predict patients' precise PTA audiometry values, mapped onto the above clinical categories. By modelling PTA as a continuous measure, we aimed to quantify impairment severity across the full range observed in the NHANES cohort (Fig. 2).

Data preprocessing

The NHANES dataset underwent preprocessing for predictive modelling. Binary features were converted to 0/1, while rejected and unknown responses were marked as missing. The dataset of 7,996 patients was split 80/20 for training and testing. Variables and patients with over 20% missing data were excluded. For the remaining missing values, we first assessed the missingness mechanism by adding missing indicators and evaluating Spearman's correlation between missing indicators and observed values. The analysis revealed no significant correlation, indicating that missing values were completely at random (MCAR) (Figure S1). Based on the findings of missing analysis, missing categorical variables were imputed using mode, and missing numerical variables were imputed using mean to preserve dataset integrity while preparing it for analysis[25–28]. The method of stratified random sampling used in this study ensured a balanced division of the 7,996 participants into training and test cohorts.
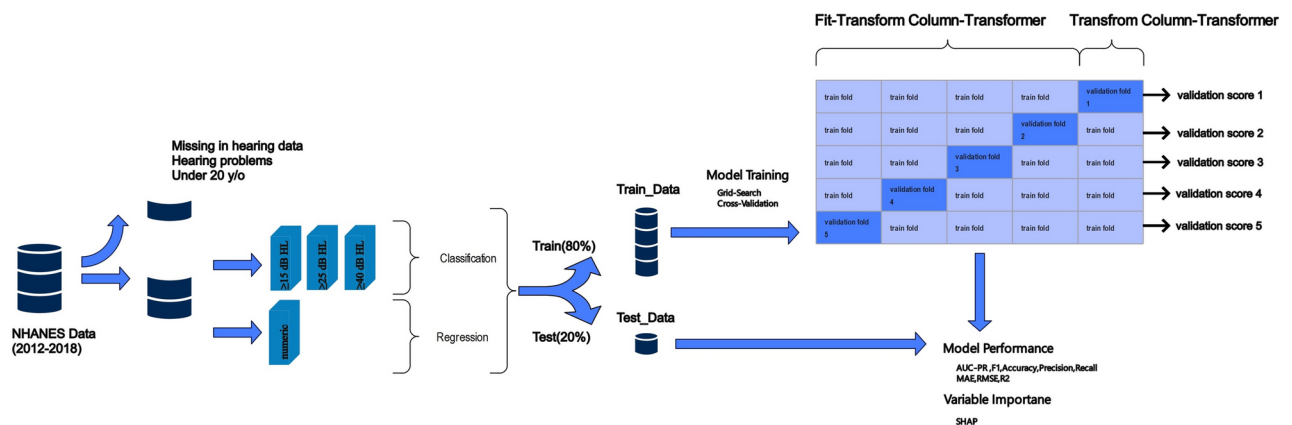

Fig. 2. Proposed methodology. Overview of the preprocessing and modeling methodology. Data preprocessing involved applying tailored transformations to each column using a column transformer, which allows numerical data to be normalized and categorical data to be encoded. The fit-transform process learns the required parameters from the training data and applies them to both training and test datasets, ensuring consistency in preprocessing. These steps prepared the data for downstream machine learning models, including custom-designed classifiers and regressors tailored for hearing loss prediction.

By creating strata based on key variables such as age, gender, and race/ethnicity, and then performing random sampling within each stratum, the approach maintained the proportional representation of these demographic features across both cohorts. To validate the comparability, chi-square tests confirmed uniform distributions for categorical variables, while continuous variables underwent Kolmogorov–Smirnov and Mann–Whitney U tests to ensure consistency in distributional shapes between the two sets. This method reduced bias and enhanced the generalizability of the predictive models developed.

Subsequent transformations included the one-hot encoding of categorical and ordinal features to convert them into a machine-readable numerical format. Numerical variables were subjected to scalar normalization to mitigate the impact of outliers, ensuring a balanced and equitable dataset for model training[29]. To address the class imbalance, present in the dataset, where the number of individuals without hearing impairment significantly outnumbered those with hearing impairment, we employed the Synthetic Minority Over-sampling Technique (SMOTE). SMOTE is a well-established oversampling approach that generates synthetic instances of the minority class by interpolating between existing minority instances and their nearest neighbors[30]. This technique has been proven effective in improving the performance of classification models on imbalanced datasets by preventing the majority class from dominating the learning process. The SMOTE algorithm was applied separately to the training set to generate synthetic minority class samples for each of the three hearing impairment thresholds (>25 dB HL, >16 dB HL, >40 dB HL). This oversampling approach ensured that the training sets used for model development were balanced, with equal representation of the positive and negative classes for each hearing impairment classification task. The number of synthetic minority class samples generated by SMOTE was determined through a grid search, optimizing for the target class distribution that yielded the best overall model performance metrics on the held-out test set.

In this study, we employed Lasso regression as a means of feature selection to enhance the predictive performance of our model and to identify the most significant predictors of hearing impairment. Lasso, known for its capacity to reduce model complexity by penalizing the absolute size of regression coefficients, can effectively zero out less relevant features, thereby facilitating a more interpretable and streamlined model. We chose Lasso given its proven efficacy in handling datasets with potential multicollinearity and ability to perform feature selection from numerous predictors. We tuned the Lasso model by setting the hyperparameter, alpha, to 0.95. This specific value was chosen to balance model complexity and predictive accuracy, ensuring that only features with substantial contribution to the prediction of hearing impairment were retained[31].

ML algorithms

To predict both the continuous PTA values and categorical hearing impairment thresholds, we evaluated an array of supervised ML algorithms. For regression modelling of the precise PTA outcome, we benchmarked linear models (Linear Regression), tree-based ensembles (Random Forest, Gradient Boosting, XGBoost, LightGBM), and multi-layer feedforward neural networks (MLP). Additionally, we employed a Multi-Layer Neural Network (MLNN), the architecture of which is detailed in a subsequent section. Classifiers explored for the dichotomous impairment assessment included logistic regression alongside the ensemble methods Random Forest, XGBoost, Gradient Boosting, CatBoost, and LightGBM[32–34].

Our models were constructed using Python (version 3.9.12; Python Software Foundation) and a collection of Python libraries, specifically numpy, pandas, matplotlib, sklearn, imblearn, xgboost, Keras, lightgbm, and catboost. The ensemble of models comprised RandomForestClassifier, RandomForestRegressor, XGBoost's XGBClassifier, XGBRegressor, GradientBoostingClassifier, GradientBoostingRegressor, LogisticRegression, LinearRegression, MLPClassifier, MLPRegressor, LightGBM's LGBMClassifier, LGBMRegressor, shap, and CatBoostClassifier (Table 1)[35,36].

MLNN architecture

We developed a multi-layer feedforward neural network for predicting PTA audiometry using the Keras API with TensorFlow 2.0 backend. The network comprised an input layer, an output layer, and four dense layers containing 32, 32, 16, and 16 nodes respectively. Rectified linear unit (ReLU) activation was applied on the hidden layers, while the output layer had a linear activation function for regressing the continuous PTA value. This choice was determined as part of the hyperparameter optimization process, as detailed in Table S2.

Category	Components
Base Environment	Python 3.9.12
Core Libraries	Numpy, Pandas, Matplotlib, Sklearn, Imblearn
Boosting Frameworks	XGBoost, LightGBM, CatBoost
Neural Network	Keras
Model Types	- RandomForest (Classifier & Regressor) - XGBoost (Classifier & Regressor) - GradientBoosting (Classifier & Regressor) - Logistic & Linear Regression - MLP (Classifier & Regressor) - LightGBM (Classifier & Regressor) - CatBoost Classifier
Interpretability	SHAP

Table 1. Technical stack and model components overview.

Dropout layers with a rate of 0.2 were used after the first and third dense layers to introduce regularization and improve generalization performance. We compiled the model using mean absolute error (MAE) loss, Adaptive Moment Estimation (Adam) optimization, and a batch size of 32.

This configuration was trained for 100 epochs while tracking performance on a held-out validation split. Using early stopping, training concluded early at 15 epochs upon observing a plateau in validation loss improvement. The final MLNN comprised 8,100 tunable parameters, achieving optimal generalizability to unseen data. Our multi-layer feedforward network for predicting continuous PTA values is outlined in Fig. 3. It depicts the propagation of input data x through four dense hidden layers, applying rectified linear activations and dropout regulation, before culminating in a linear output layer to regress the PTA outcome ŷ.

Evaluation of model performance

We evaluated model performance using several metrics tailored to the regression and classification tasks, accounting for the imbalanced nature of our dataset.

For the regression models predicting continuous PTA, we focused on mean absolute error (MAE), and root mean squared error (RMSE). These metrics provided a robust framework for quantifying the deviation of our model predictions from the actual values, thus offering a clear picture of model accuracy and the consistency of its predictions. RMSE is particularly useful in highlighting the impact of outliers on model performance, while MAE offers a straightforward, interpretable measure of average error magnitude across predictions. Lower values indicate better model fit and prediction accuracy.

The classification models used the speech-frequency pure-tone average (PTA) thresholds as the ground truth for determining hearing impairment categories. These thresholds (> 25 dB HL, > 16 dB HL, and > 40 dB HL), previously defined in the Primary Outcome subsection, were applied to categorize participants into distinct hearing impairment groups for the classification tasks. For our classification models based on PTA thresholds, accuracy alone was an insufficient metric given the class imbalance. In addition to accuracy, we considered the F1-score, the area under the receiver operating characteristic curve (AUCROC), and the area under the precision-recall curve (AUPRC). AUCROC and AUPRC provided valuable insights into the models' ability to discriminate between classes, with AUPRC being particularly useful for evaluating the performance of the minority class[37]. We also examined specificity, which measured the proportion of true negatives correctly identified, and sensitivity, which quantified the true positive rate. The F1-score balanced precision and recall, emphasizing minimizing both false positives and false negatives – an important consideration for this clinical prediction task[38].

We utilized stratified fivefold cross-validation repeated across 5 iterations on the NHANES training set to derive stable estimates of the above metrics (Fig. 2), including 95% confidence intervals. This approach involved partitioning the training data into five stratified folds, with each iteration reshuffling and redistributing the data. The 95% CIs for model performance metrics were calculated using the variability observed across the repeated fivefold cross-validation iterations. The optimal hyperparameters yielding peak average test AUPRC across validation folds were selected as it is particularly sensitive to class imbalance, focusing on the minority class due to original nature of the test dataset (Table S2).

It is important to note that the evaluation of model performance was conducted on a hold-out test dataset that was separated prior to model training to prevent data leakage and ensure an unbiased estimate of performance. While this approach offers robust insights into the models' predictive accuracy within the NHANES dataset, it does not constitute an independent validation using a completely separate dataset from a different source or population.

Pairwise comparisons using Nemenyi's test were conducted to identify specific performance differences between individual models.
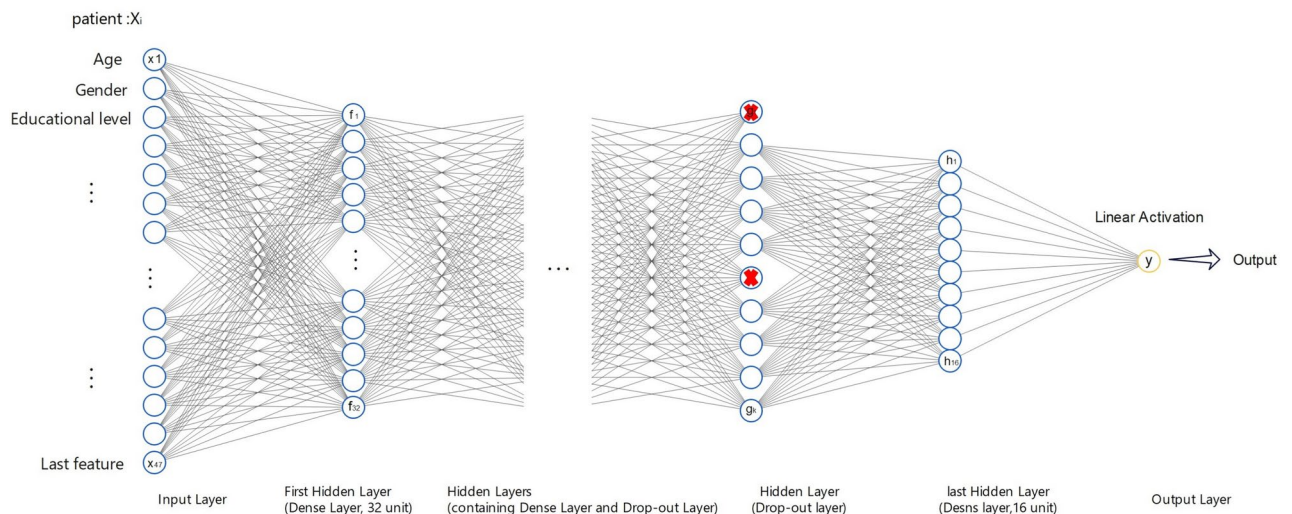

Fig. 3. Overview of our MLNN.

Model interpretation

To interpret the influential predictors in our top-performing models, we utilized SHapley Additive exPlanations (SHAP). SHAP analysis assigns an importance value to each feature based on its contribution to the prediction, both at the individual sample level and across the entire dataset[39].

For our optimal regression model, we generated a SHAP summary plot consolidating feature importance rankings with visualizations of each predictor's effect magnitude and direction. This allowed us to identify the subset of variables exerting the largest positive and negative impacts on predicted PTA.

Our classification framework prioritized SHAP values specific to the positive class-identifying impairment. We plotted SHAP feature importance and visualized SHAP value distributions comparing impaired patients to the overall cohort using SHAP summary and cohort bar plots. This highlighted the key risk factors differentially influencing hearing impairment predictions.

In addition to global perspectives, we inspected local SHAP explanations for individual patients across various demographic and clinical profiles. This localized SHAP interpretation uncovered patterns of predictor contributions specific to subjects matched on characteristics like age, sex, and comorbidities[40].

Result

Sample characteristics

This retrospective study analyzed hearing threshold data from 7,996 participants collected between 2012 and 2018 (Fig. 1). The participants were randomly divided into training and test cohorts that were well-matched in terms of demographic and baseline characteristics. Specifically, as shown in Table 2, the mean age was 44.19 ± 14.28 years, with 50.02% of the participants being male. The majority (73.78%) exhibited normal hearing thresholds of 16 dB or less.

Performance in identifying the hearing status

We employed an extended feature ML model to predict three severity tiers, as outlined in the methodology section, within a binary framework. Our analysis ensured that none of the models exhibited overfitting, as evidenced by the close alignment of accuracies between the training and test sets. We focused on classifying hearing impairment based on thresholds greater than 25 dB HL to emphasize the importance of early detection and management in preventing progression. Table 3 presents an overview of the ML models' performance in the classification setting utilizing the > 25 dB HL threshold. Across all models, the accuracy, AUC-ROC, and AUPRC exceeded 75%, 0.79, and 0.69, respectively. Notably, the LightGBM model emerged as the most effective overall, achieving an accuracy of 80.1% and 79.8% for the training and test groups, respectively, along with corresponding AUC-ROC values of 0.809 and 0.820 and competitive AUPRC values of 0.738 and 0.729. CatBoost and Gradient Boosting are the next best performers, with test set accuracies around 0.79 and F1-scores around 0.79–0.80. The MLP and Logistic Regression models show slightly lower, but still reasonably good, test set accuracies of 0.784 and 0.788 respectively, with F1-scores around 0.78–0.79. In contrast, XGBoost has the lowest test set accuracy of 0.756 and F1-score of 0.778 among the models evaluated. The F1-scores, presented in Tables 3, demonstrated that the LightGBM model achieved the highest balance between precision and recall across all classification thresholds. In classifying > 25 dB HL thresholds, LightGBM attained an F1-score of 0.800 in the test dataset, reflecting its robustness in minimizing both false positives and false negatives in this clinical prediction context.

All ML models exhibited an accuracy, AUCROC, and AUPRC above 73%, 0.870, and 0.707 in predicting HL thresholds above 16 dB (Table S3). Among these, LightGBM emerged as the optimal model for > 16 dB HL classification, achieving an accuracy of 77.6%, an AUC-ROC of 0.891, and an AUPRC of 0.786 for the training cohort, and 78.8% accuracy, 0.889 AUC-ROC, and 0.791 AUPRC for the test cohort. The Random Forest and XGBoost models also achieved strong results, with test set accuracies of 0.771 and 0.784, respectively. The Gradient Boosting and Logistic Regression models performed reasonably well, while the multilayer perceptron (MLP) neural network had the lowest performance among the evaluated models.

In the classification task targeting HL thresholds above 40 dB, all models achieved an accuracy of over 85.1%, an AUC-ROC of 0.761, and an AUPRC of 0.634 (Table S4). Notably, the LightGBM model outperformed others in this context, achieving a remarkable accuracy of 87.3% for the training group and 86.1% for the test group, along with corresponding AUC-ROC values of 0.792 and 0.789 and AUPRC values of 0.730 and 0.721, respectively. The XGBoost, Random Forest, and Logistic Regression models also demonstrated solid performance, with test set accuracies ranging from 0.856 to 0.871. Interestingly, the MLP model showed a slight improvement in its test set accuracy (0.871) compared to the HL > 16 dB task, narrowing the gap with the top-performing models.

Figure 4 illustrates the predictive capability of all ML models in classifying > 25 dB HL, with the evaluation metric AUC-ROC being employed to assess the model's sensitivity–specificity balance. Higher AUC-ROC values indicate superior performance of the model across various ML models.

For a comprehensive comparison of model performances across different classification scenarios, Fig. 5 and S2 present a detailed analysis including Random Forest, XGBoost, Gradient Boosting, MLP, logistic regression, LightGBM, and CatBoost models.

Performance in predicting hearing thresholds

We conducted an evaluation of regression models to gauge their efficacy in predicting hearing thresholds. Lower MAE and RMSE values, along with higher R2 scores, indicate stronger predictive performance. Our analysis encompassed eight widely utilized ML models, including Random Forest, XGBoost, Gradient Boosting, MLP, logistic regression, LightGBM, CatBoost, and MLNN. Figure 6 provides a concise summary of their performance metrics derived from five-fold cross-validation.

To address the potential for overfitting, we implemented several strategies. Firstly, we utilized dropout layers, which randomly deactivate a proportion of the neurons during training, helping to prevent the model from

Variables	Participants (n = 7996)	Training cohort (n = 6396)	Test cohort (n = 1600)
Age (year), mean (SD)	44.19 ± 14.28	44.21 ± 14.29	44.14 ± 14.28
Gender, n(%)	7996 (100)	6396 (100)	1600 (100)
female	3999 (50.02)	3204 (50.09)	795 (49.68)
male	3997 (49.98)	3192 (49.91)	805 (50.32)
Race, n(%)	7996 (100)	6396 (100)	1600 (100)
Mexican American	1163 (14.55)	934 (14.60)	229 (14.31)
Other Hispanic	949 (11.86)	766 (11.97)	183 (11.43)
Non-Hispanic White	2582 (32.3)	2069 (32.35)	513 (32.07)
Non-Hispanic Black	1970 (24.64)	1577 (24.66)	393 (24.56)
Other Race—Including Multi-Racial	1332 (16.66)	1050 (16.42)	282 (17.63)
Education, n(%)	7996 (100)	6396 (100)	1600 (100)
Less than high school degree	1652 (20.66)	1317 (20.59)	335 (20.94)
High school graduate or some college degree	1713 (21.41)	1363 (21.31)	350 (21.87)
College graduate or above	4631 (57.92)	3716 (58.10)	915 (57.19)
Cotinine level (ng/ml), mean(SD)	56.40 (24.58)	56.23 (24.66)	57.1 (25.61)
Smoking, n(%)	7986 (100)	6389 (100)	1597 (100)
Current or Former smoker[1]	3329 (41.69)	2645 (41.40)	684 (42.83)
Never	4657 (58.31)	3744 (58.60)	913 (57.17)
Diabetes Mellitus, n(%)	7991 (100)	6391 (100)	1600 (100)
Yes	933 (11.68)	749 (11.72)	184 (11.50)
No	6893 (86.26)	5511 (86.24)	1382 (86.37)
Borderline	165 (2.06)	131 (2.04)	34 (2.13)
Hypertension, n(%)	7989 (100)	6393 (100)	1596 (100)
Yes	2476 (30.99)	1975 (30.89)	501 (31.39)
No	5513 (69.01)	4418 (69.11)	1095 (68.61)
BMI (kg/m^2), mean (SD)	29.40 (7.20)	29.55 (7.25)	29.11 (7.09)
Height (m), mean (SD)	167.53 (10.06)	167.44 (9.98)	167.71 (10.11)
Weight (kg), mean (SD)	82.71 (22.17)	82.82 (22.11)	82.52 (21.89)
HDL (mg/dl), mean (SD)	52.73 (15.99)	52.75 (16.01)	52.69 (15.90)
LDL (mg/dl), mean (SD)	114.53 (35.25)	114.50 (35.21)	114.62 (35.30)
Total Cholesterol (mg/dl), mean (SD)	192.67 (41.32)	192.65 (41.31)	192.71 (41.33)
Triglyceride (mg/dl), mean (SD)	156.35 (138.55)	156.41 (138.60)	156.14 (138.30)
Hyperlipidemia, n(%)	7951 (100)	6376 (100)	1575 (100)
Yes	2440 (30.69)	1958 (30.71)	482 (30.60)
No	5511 (69.31)	4418 (69.29)	1093 (69.40)
Vigorous activities (minutes), mean (SD)	26.46 (48.90)	26.34 (48.58)	26.81 (49.10)
Moderate activities (minutes), mean (SD)	36.80 (54.96)	36.77 (54.91)	36.86 (55.12)
Congestive heart failure, n(%)	7981 (100)	6386 (100)	1595 (100)
Yes	158 (1.98)	128 (2.01)	30 (1.88)
No	7823 (98.02)	6258 (97.99)	1565 (98.12)
Coronary heart disease, n(%)	7971 (100)	6377 (100)	1594 (100)
Yes	169 (2.12)	140 (2.19)	29 (1.82)
No	7802 (97.88)	6237 (97.81)	1565 (98.18)
History of angina, n(%)	7981 (100)	6389 (100)	1592 (100)
Yes	128 (1.60)	101 (1.58)	27 (1.70)
No	7853 (98.40)	6288 (98.42)	1565 (98.30)
Heart attack, n(%)	7990 (100)	6395 (100)	1595 (100)
Yes	193 (2.42)	157 (2.46)	36 (2.26)
No	7797 (97.58)	6238 (97.54)	1559 (97.74)
Stroke, n(%)	7991 (100)	6392 (100)	1599 (100)
Yes	196 (2.45)	161 (2.52)	35 (2.19)
No	7795 (97.55)	6231 (97.48)	1564 (97.81)
Protein intake (gm), mean (SD)	84.07 (42.92)	84.11 (42.94)	83.95 (42.86)
Fiber intake (gm), mean (SD)	17.43 (11.22)	17.39 (11.20)	17.55 (11.31)
Potassium intake (mg), mean (SD)	2648.74 (1272.99)	2647.98 (1272.68)	2652.64 (1273.52)
Cholesterol intake (mg), mean (SD)	307.45 (246.64)	308.01 (246.89)	306.32 (245.89)
Continued			

Variables	Participants (n = 7996)	Training cohort (n = 6396)	Test cohort (n = 1600)
Saturated fat intake (mg), mean (SD)	26.65 (16.97)	26.71 (17.02)	26.12 (16.48)
Total fat intake (gm), mean (SD)	83.11 (47.31)	83.07 (47.29)	83.78 (47.48)
Magnesium intake (mg), mean (SD)	305.60 (156.12)	305.57 (156.11)	305.73 (156.16)
Calcium intake (mg), mean (SD)	933.05 (584.21)	933.12 (584.27)	932.72 (584.13)
Sodium intake (mg), mean (SD)	3644.83 (1871.31)	3643.98 (1870.78)	3646.23 (1872.01)
Kilocalories intake, mean (SD)	2180.96 (999.81)	2180.87 (999.78)	2181.12 (999.98)
Hearing threshold categories, n (%)	7996 (100)	6396 (100)	1600 (100)
Normal (< 16 dB)	5899 (73.78)	4708 (73.61)	1191 (74.44)
Slight (16-25db)	1379 (17.25)	1108 (17.32)	271 (16.94)
Mild (26–40 dB)	569 (7.12)	455 (7.11)	114 (7.13)
Moderate or worse (> 40)	149 (1.86)	125 (1.95)	24 (1.50)
Pure-tone average, mean (SD)	10.54 (9.80)	10.71 (9.86)	9.87 (8.64)

Table 2. Baseline demographic and clinical characteristics. 1 at least 100 cigarettes in life.

Model		Accuracy	Precision	Recall (sensitivity)	F1 score	AUPRC	AUCROC	Specificity	NPV
LightGBM	Train set	0.801 (0.800–0.801)	0.830 (0.828–0.832)	0.773 (0.769–0.776)	0.800 (0.799–0.802)	0.738 (0.738–0.738)	0.809 (0.808–0.810)	0.742 (0.740–0.745)	0.971 (0.970–0.971)
	Test set	0.798	0.841	0.762	0.800	0.729	0.820	0.731	0.969
RF	Train set	0.796 (0.796–0.796)	0.823 (0.822–0.825)	0.751 (0.749–0.754)	0.818 (0.815–0.820)	0.785 (0.784–0.788)	0.801 (0.801–0.802)	0.734 (0.733–0.736)	0.967 (0.965–0.968)
	Test set	0.799	0.828	0.769	0.797	0.731	0.801	0.721	0.969
CatBoost	Train set	0.781 (0.780–0.783)	0.833 (0.831–0.836)	0.775 (0.772–0.777)	0.802 (0.799–0.804)	0.709 (0.707–0.710)	0.801 (0.800–0.803)	0.755 (0.754–0.757)	0.971 (0.968–0.973)
	Test set	0.792	0.843	0.749	0.790	0.712	0.807	0.728	0.967
MLP	Train set	0.773 (0.772–0.774)	0.832 (0.830–0.833)	0.707 (0.705–0.708)	0.764 (0.762–0.765)	0.698 (0.689–0.704)	0.800 (0.798–0.801)	0.742 (0.742–0.742)	0.962 (0.959–0.964)
	Test set	0.784	0.853	0.721	0.781	0.701	0.803	0.741	0.964
Gradient Boosting	Train set	0.771 (0.770–0.773)	0.849 (0.841–0.841)	0.759 (0.758–0.761)	0.801 (0.800–0.801)	0.719 (0.718–0.721)	0.808 (0.808–0.808)	0.771 (0.769–0.775)	0.970 (0.966–0.973)
	Test set	0.777	0.866	0.748	0.803	0.709	0.791	0.766	0.968
Logistic Regression	Train set	0.769 (0.768–0.771)	0.827 (0.826–0.829)	0.742 (0.740–0.743)	0.782 (0.781–0.782)	0.721 (0.720–0.723)	0.805 (0.804–0.807)	0.733 (0.731–0.734)	0.966 (0.965–0.969)
	Test set	0.788	0.852	0.735	0.789	0.719	0.790	0.721	0.965
XGBoost	Train set	0.768 (0.767–0.768)	0.812 (0.811–0.814)	0.747 (0.745–0.748)	0.778 (0.776–0.779)	0.717 (0.716–0.719)	0.798 (0.789–0.803)	0.699 (0.688–0.710)	0.965 (0.964–0.966)
	Test set	0.756	0.828	0.734	0.778	0.723	0.776	0.689	0.963

Table 3. ML models' performance in HL threshold greater than 25.

relying too heavily on specific features and improving its generalization capability. Additionally, we employed early stopping, whereby the training process was halted when the validation loss stopped improving, ensuring the model did not overfit to the training data. Furthermore, the close alignment of the training and test set metrics suggests the MLNN model was able to generalize well to unseen data.

Notably, our findings revealed consistent performance across most models, as evidenced by nearly equivalent mean MAE values between the training and test datasets, suggesting a lack of overfitting (Table 4). MLNN outperformed all other models with a significantly better performance (p < 0.001). However, Linear Regression had significantly poorer performance across all metrics (p < 0.01), indicating potential limitations in its predictive capabilities. Statistical analysis revealed that six other models had comparable performance levels (p = 0.42).

The MLNN established the highest predictive capability, as evidenced by its lowest MAE of 3.05 dB (95% CI: 3.04 to 3.06) and RMSE of 4.49 dB (95% CI: 4.48 to 4.50), along with the highest R2 of 0.69. Among other ML models assessed, Gradient Boosting emerged as the top performer, exhibiting the most favorable performance metrics regarding predictive accuracy. Specifically, it achieved the lowest MAE of 5.54 dB (95% CI: 5.38 to 5.69) and RMSE of 8.28 dB (95% CI: 8.10 to 8.47). Figure 7 shows the learning curve associated with MLNN, illustrating its convergence to optimal performance through iterative cross-validation. Notably, the curve
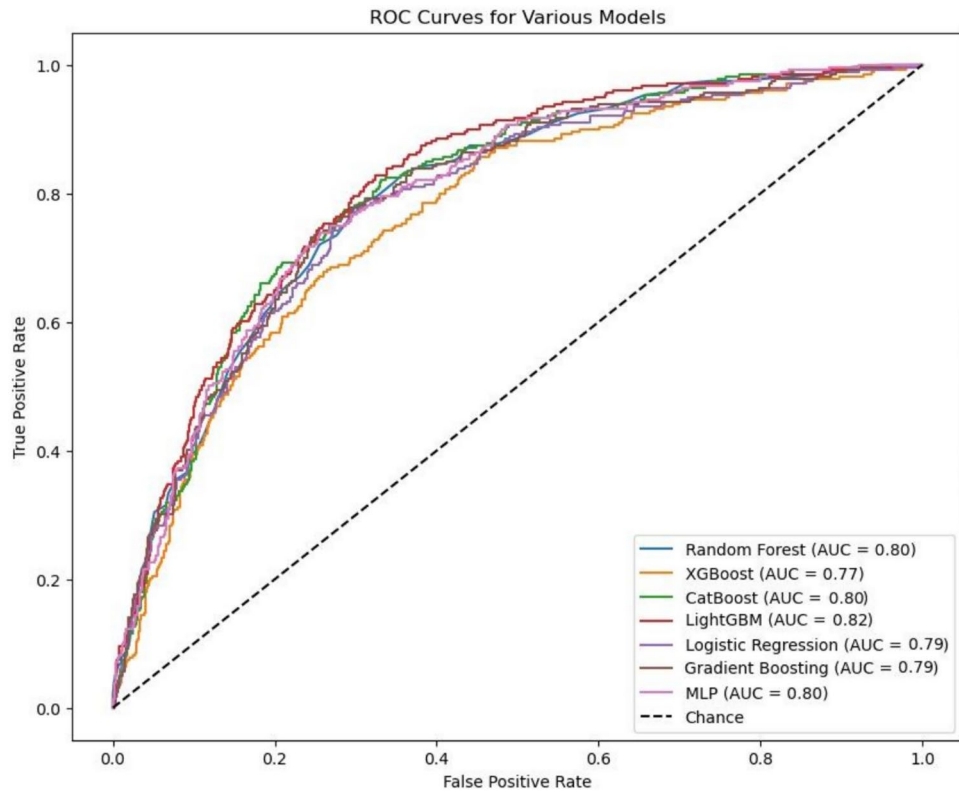
Fig. 4. AUC-ROC curve of LightGBM model for > 25 dB HL classification.
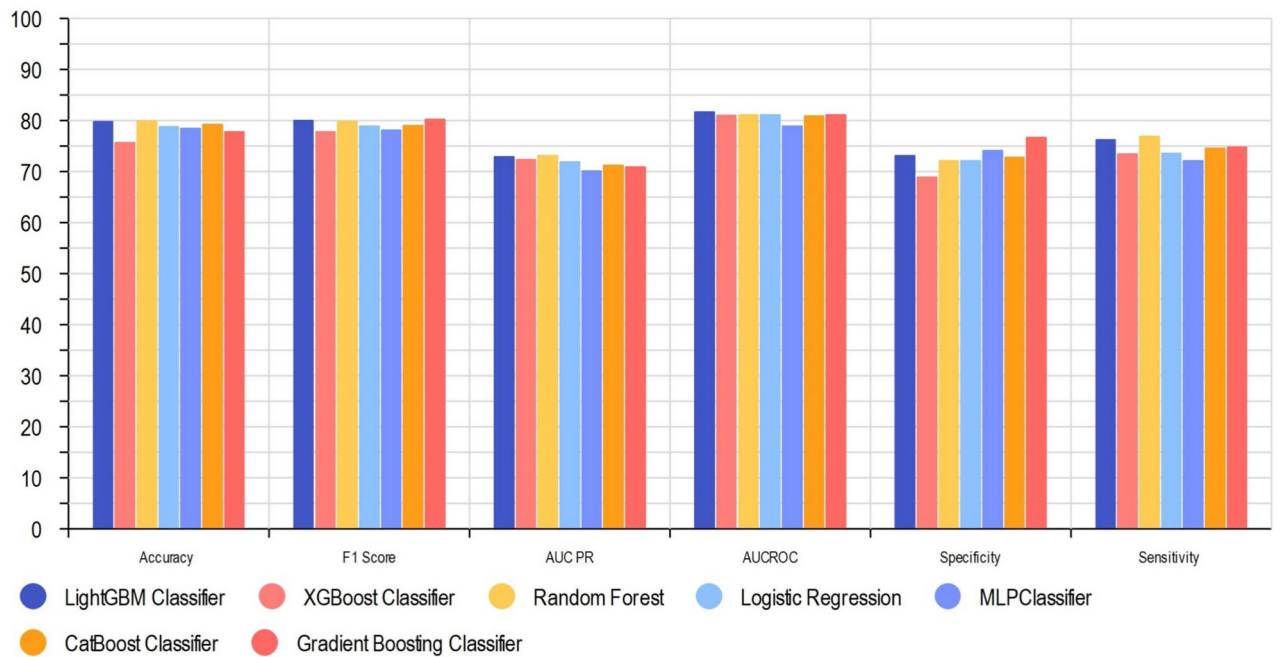

Fig. 5. Performance comparison of ML models for > 25 dB HL classification.

demonstrates that the MLNN model achieved its peak performance following approximately 100 iterations of five-fold cross-validation. This observation underscores the effectiveness of MLNN in learning complex patterns within the data and attaining superior predictive accuracy.
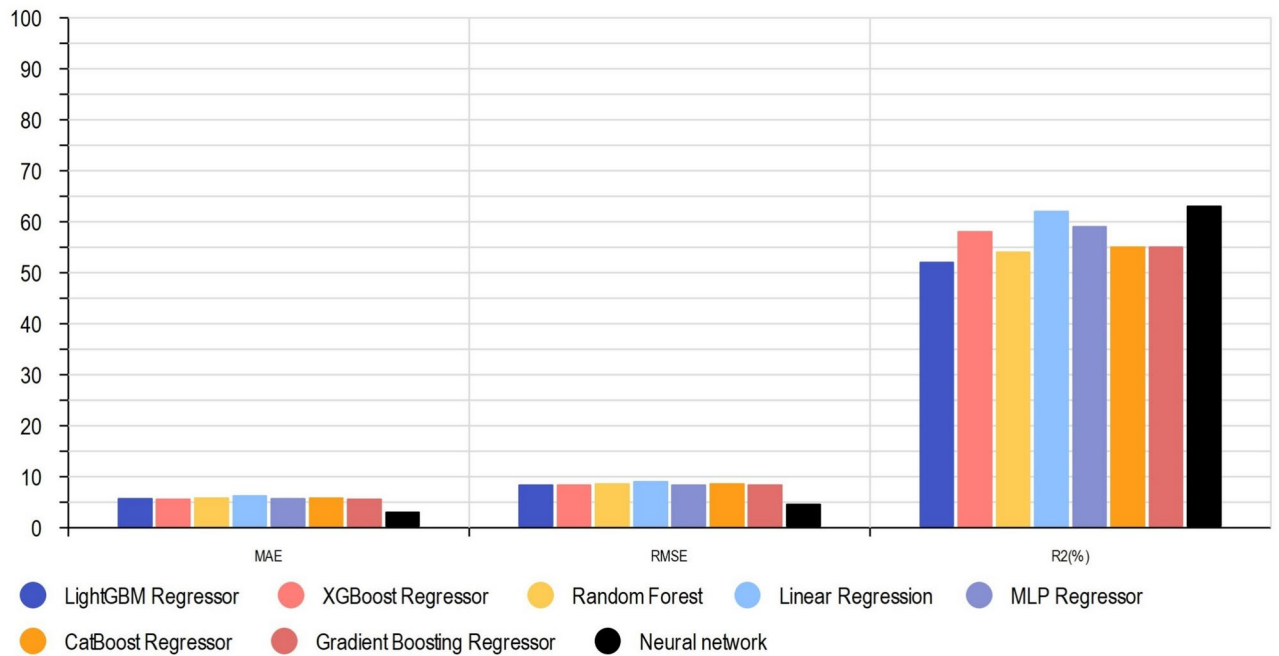
Fig. 6. ML models' performance of ML models for predicting hearing thresholds.

Model		MAE	RMSE	R² (%)
Neural Network	Train set	3.05 (3.04–3.07)	4.49 (4.46–4.51)	69
	Test set	3.06	4.52	63
Gradient Boosting	Train set	5.54 (5.52–5.55)	8.28 (8.13–8.36)	60
	Test set	5.60	8.31	55
XGBoost	Train set	5.54 (5.51–5.56)	8.37 (8.36–8.40)	56
	Test set	5.61	8.38	58
MLP	Train set	5.61 (5.61–5.61)	8.30 (8.28–8.33)	58
	Test set	5.63	8.34	59
LightGBM	Train set	5.66 (5.64–5.67)	8.51 (8.49–8.52)	56
	Test set	5.68	8.55	57
CatBoost	Train set	5.75 (5.74–5.77)	8.54 (8.54–8.54)	56
	Test set	5.77	8.57	55
RF	Train set	5.80 (5.79–5.80)	8.51 (8.49–8.52)	56
	Test set	5.77	8.57	54
Linear Regression	Train set	6.05 (6.03–6.06)	8.96 (8.95–8.98)	53
	Test set	6.20	8.98	52

Table 4. ML models' performance in prediction of hearing threshold.

Feature importance

We selected the most important variables of the basic features for the development of our ML models, choosing those with the highest performance based on LASSO feature selection. All variables, along with their corresponding names in the NHANES dataset, are outlined in Table S1 for clarity and reference. To gain insights into the impact of each variable on predicting hearing status and thresholds, we utilized SHAP values. Figure 8 visually presents the magnitude and direction of influence exerted by each factor on the outcome, as determined by the Gradient Boosting model. This graphical representation demonstrates how individual variables affect the predictive results. Each sample is represented as a point per feature and the x-axis shows the feature's effect. The red segments of each graph indicate higher values for that feature in the test set data, relating to an increased probability of the model predicting an individual as a patient.

In the classification task targeting HL thresholds exceeding 25 dB, our analysis revealed age as the most influential predictor, followed by gender and systolic blood pressure. Specifically, younger age, female gender, and lower systolic blood pressure were associated with a higher likelihood of falling below the 25 dB HL threshold (Fig. 8a). Similarly, in predicting hearing thresholds, age emerged as the predominant predictor, followed by
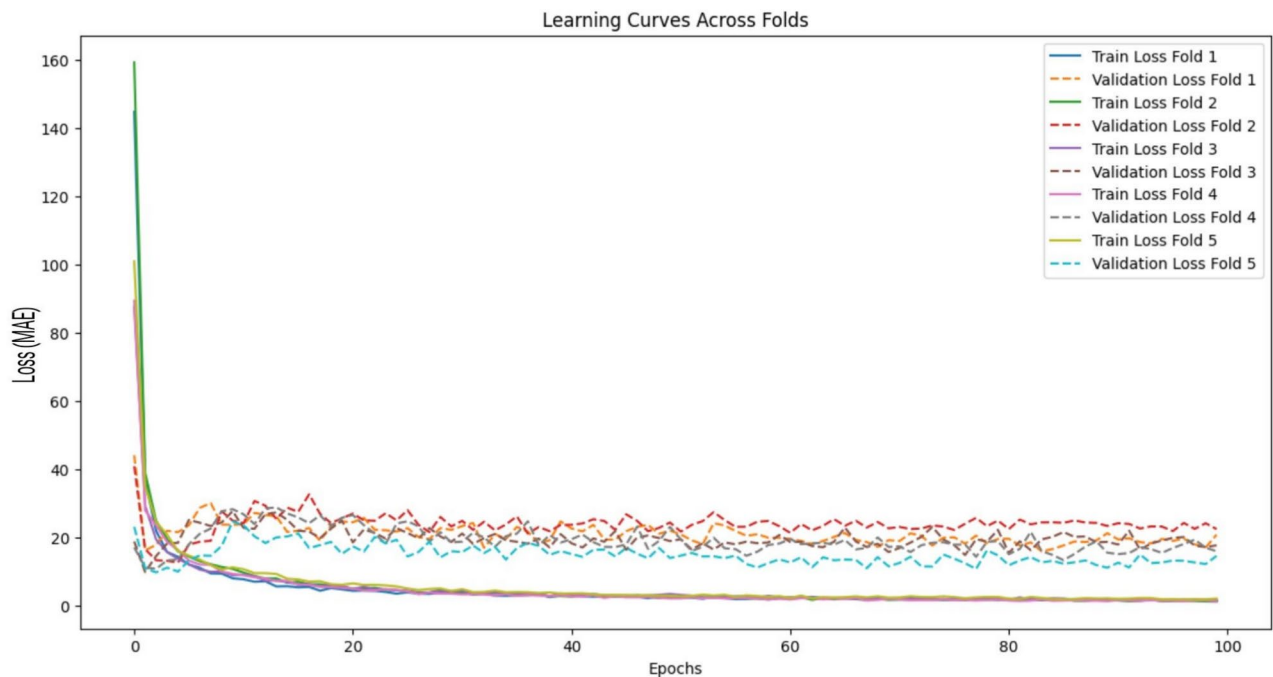
Fig. 7. The learning curve in the MLNN.

gender and education (Fig. 8b). Notably, age exhibited significantly higher SHAP values for both predicting hearing status and thresholds, indicating its pivotal role in the model's predictions. The SHAP also highlights the importance of diabetes in understanding hearing outcomes with a value greater than 1.

To provide an overview of the model's predictive power, we visualized the feature importance ranking based on SHAP values in Figs. 9a and 9b. Figures 9a and 9b depict SHAP-derived feature importance rankings for the classification (> 25 dB HL threshold) and regression (continuous PTA values) tasks, respectively.

We conducted a sensitivity analysis by excluding age and gender from our feature set to assess model robustness, as detailed in Figure S3 of the supplementary materials.

We employed SHAP individual force plots to provide a visual representation of how each feature contributes to the model's decision-making process, thereby enhancing our understanding of the predictive mechanisms involved. These plots arrange features from left to right, with positive contributions depicted on the left and negative contributions on the right. In Fig 10a and b , we present SHAP individual force plots for two participants, labeled as number 1 and 47 in the case of participant number 1, the model's output value [f(x)] for the classification of HL thresholds exceeding 25 dB is −2.42. This negative value indicates the model's prediction that the patient falls below the 25 dB HL threshold. To derive the probability from this output value, we employed the following formula.

$$p = \frac{1}{1 + e^{-(output(f(x)) - basevalue)}}$$

where 'e' represents Euler's number, which is approximately equal to 2.71828. This calculation enables us to interpret the model's output in terms of the likelihood of the patient belonging to a specific HL threshold category.

The calculated probability of approximately 0.28 indicates a 28% chance of the participant belonging to the positive group. A probability below 0.5 suggests that the model predicts the participant as a member of the negative class. Notably, factors such as age and male gender prominently influence the prediction of a lower likelihood of a hearing threshold exceeding 25 dB (refer to Fig. 10a). Conversely, in the case of participant number 47, a low waist circumference exerts a positive influence, while age, male gender, high sodium intake, and elevated levels of triglycerides and systolic blood pressure contribute negatively to the outcome, resulting in a f(x) of 0.81 (see Fig. 10b). With a positive output value and a calculated probability of 0.90, the model confidently predicts the participant as a member of the positive class. While these SHAP values demonstrate statistical associations between patient characteristics and hearing loss predictions, they should not be interpreted as causal relationships for clinical intervention, as establishing causality would require additional targeted clinical studies.

Discussion

In this study, we developed a risk prediction system by utilizing the best-performing ML models on a representative sample of US adults from the NHANES dataset between 2012 and 2018. These models were scrutinized for their effectiveness in two specific areas: the classification of HL into three categories (> 25 dB HL, > 16 dB HL, and > 40 dB) and the accurate prediction of the PTA threshold level. The outcomes indicated that all models achieved at least 75% accuracy in in the main classification with the LightGBM outperforming the other
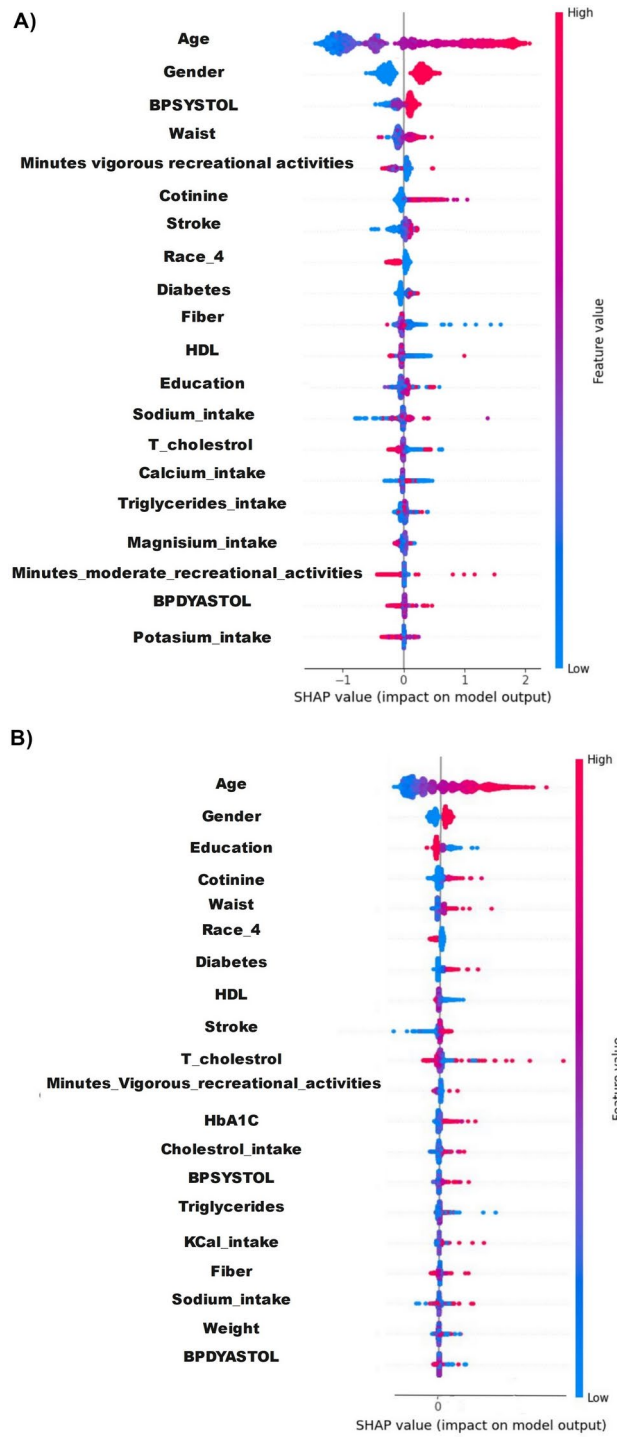
Fig. 8. SHAP values. **A)** SHAP values in > 25 dB HL threshold classification, **B)** SHAP values in predicting hearing threshold.

models in terms of predictive Accuracy, AUCROC, and AUPRC. Also, the MLNN comparatively outperformed all models concerning the metrics of MAE, RMSE, and R^2. The analysis identified primary associated factors with significant influence on hearing outcome: advanced age, male gender, elevated systolic blood pressure, increased waist circumference, reduced participation in vigorous recreational activities, smoking, stroke, race, and diabetes. Among the performance metrics, the F1-score was particularly relevant in assessing the balance between precision and recall, which is crucial for clinical tasks where both false positives and false negatives carry significant implications. The LightGBM model's superior F1-scores across all thresholds underscore its capacity to achieve this balance, further validating its clinical utility for early identification of hearing impairment.
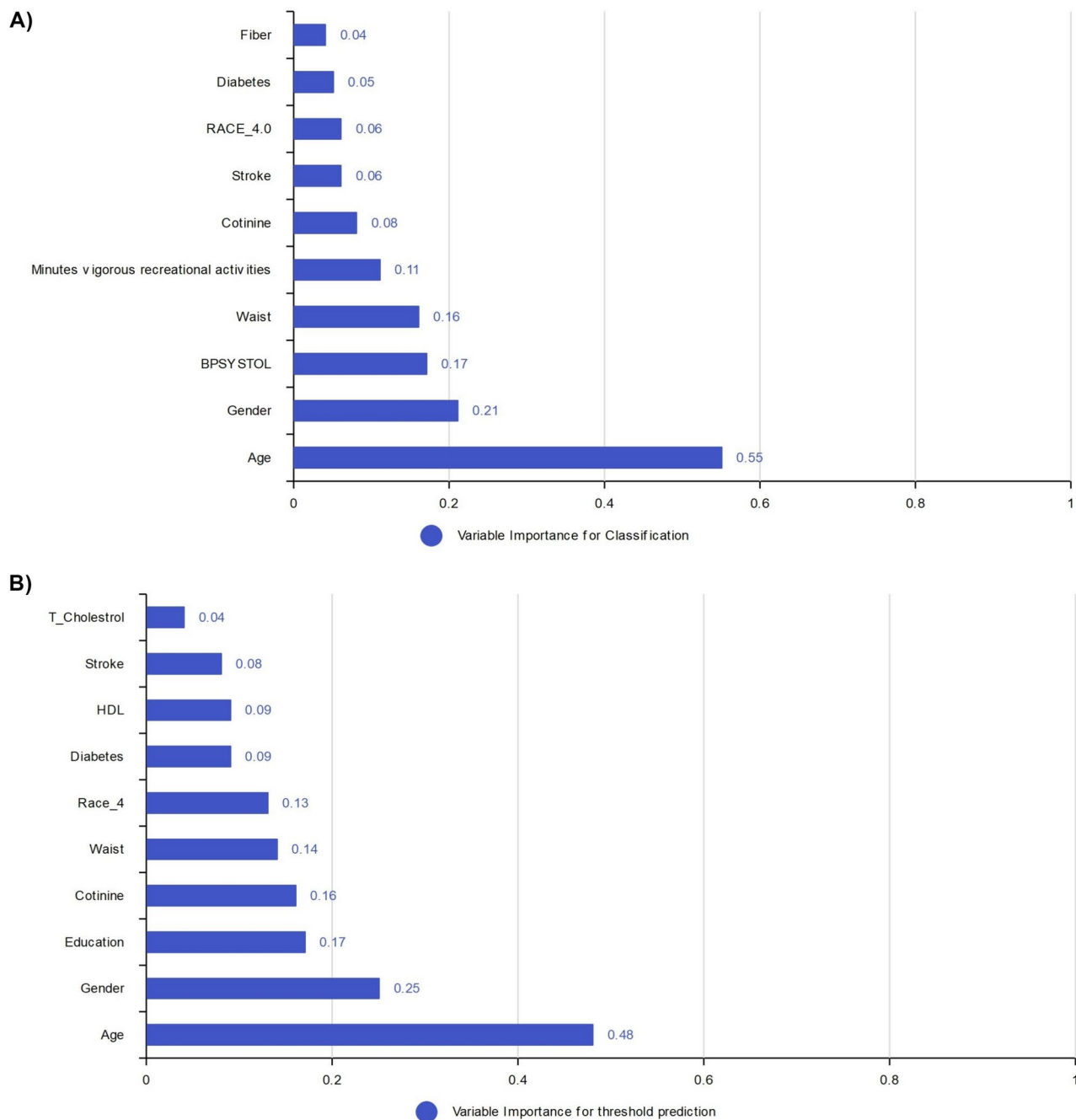
Fig. 9. Feature importance ranking. **A)** Feature importance ranking in > 25 dB HL threshold classification, **B)** Feature importance ranking in predicting hearing threshold.

ML has found numerous applications in audiology and is believed to have higher accuracies and lower prediction errors than regression models[19]. One such application is the prediction of hearing loss in workers exposed to high levels of noise in industrial settings[41]. Additionally, ML has been successfully employed in the classification of auditory brainstem response and audiogram, yielding promising results[18,42]. Previous ML models for predicting hearing outcomes were limited by small sample sizes, lack of transparency, few models examined, and a narrow consideration of risk factors including demographics, medical history, military background, noise exposure, and self-reported hearing capabilities[19,24,43]. Ellis and Souza applied ML in forecasting categorical HL to assess the slope between 2 and 4 kHz pure-tone thresholds using three classifiers[43]. Among these, the RF algorithm was identified as the most effective for the utilization of limited audiometric and demographic information. However, the precision of these results was somewhat limited, with only 55% of individuals accurately categorized for categorical hearing outcomes based on limited input features. Impressively, in our investigation, the CVD risk factor-based LightGBM model increased accuracy for the training and testing cohorts in classifying HL categories. LightGBM offered the best performance in our experience which was not
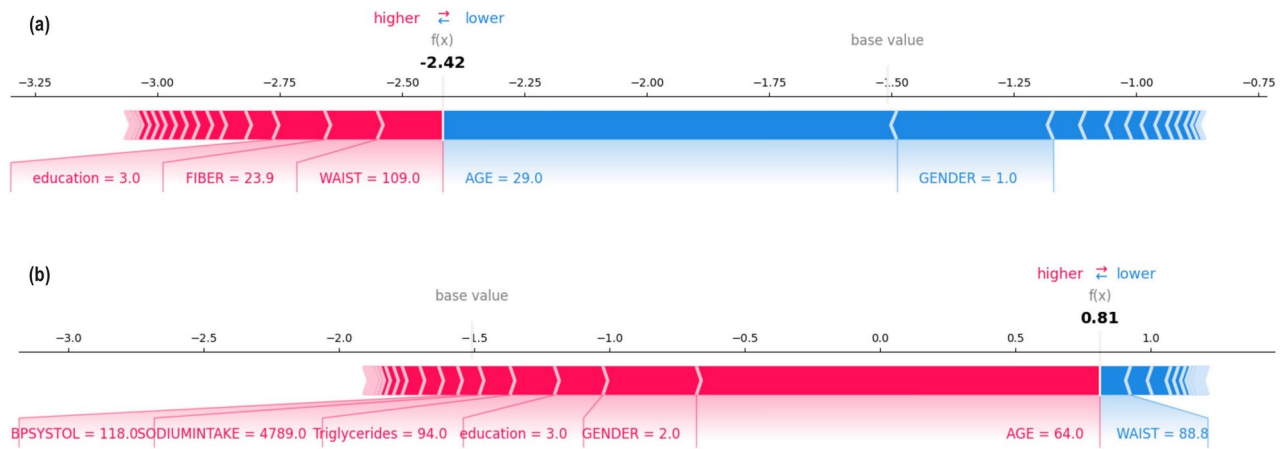
Fig. 10. SHAP individual force plot for the first and 47th participants. This visualization focuses only on the most influential factors in a patient's hearing prognosis. Less impactful attributes are omitted. Arrows depicting each featured factor point to either impaired or normal hearing, with red and blue respectively. Arrow length correlates to the feature's strength according to SHAP values. A central line denotes the patient's predicted hearing outcome probability numerically above. Arrows extend from this, labelled beneath with the feature and its value. The positioning and sizing intuitively represent how features collectively shift risk up or down. More details in the text. **A)** SHAP individual force plot for the first participant with a low possibility of hearing impairment **B)** SHAP individual force plot for 47th participant with a high possibility of hearing impairment.

unexpected. Its gradient-based learning techniques focus on reducing data instances and features during tree construction, resulting in faster training and improved generalization. Additionally, its leaf-wise tree growth strategy enables more expressive representations with fewer levels, often leading to better accuracy[44]. Also, a study conducted by Soylemez et al. on 200 workers from a metal industry utilized a support vector machine (SVM) model, which exhibited promising predictive performance with an accuracy of 90%, F1 score of 91%, precision of 95%, and recall of 88%; however, the limited sample size and consideration of only a few risk factors, such as age, noise exposure, and tinnitus, may make the findings prone to overfitting and limit the generalizability of the results[45].

The incorporation of precise PTA estimations in our study aims to serve as a complementary screening tool to identify individuals at higher risk for hearing loss. This approach can streamline referral pathways to comprehensive audiological evaluations, which remain indispensable for accurate diagnosis and holistic management of hearing health. Gathman et al. also considered features like demographics, medical history, and self-evaluated hearing to predict the exact HL[24]. Their findings indicated that the five key predictors for a higher PTA included older age, poorer self-assessed hearing, male gender, higher body mass index, and smoking history. While they solely employed the LightGBM model, our study explored eight ML models, including the custom-designed MLNN. The LightGBM model in their research achieved an MAE in the test set between predicted and actual PTA of 5.29 dB HL (95% CI: 4.97-5.61), comparable to 5.66 dB HL (95% CI: 5.48–5.84) in our study[24]. Likewise, we observed the optimal performance with the MLNN, which recorded an MAE of 3.05 dB HL (95% CI: 3.04 to 3.06), compared to 5.29 dB HL in their study. Additionally, we provided RMSE and R2 scores for a more comprehensive comparative analysis. These results suggest that the neural network is more robust to variations in input characteristics and might act as a good feature extractor in our hearing level data. As illustrated in Fig. 4, the MLNN's capability to discern intricate patterns within the dataset and achieve superior predictive accuracy was evident. The MLNN model excels in managing complex interactions and a high number of parameters, offering robustness and superior generalization compared to linear models, with linear regression as a specific subset[46].

In response to the persistent criticism of ML models due to their inability to clarify the cause-effect relationships owing to the black-box nature of ML algorithms, we have incorporated explainability and interpretability characteristics into our LightGBM model by employing the SHAP technique. The SHAP Framework is designed to elucidate the correlations that ML predictive models identify by highlighting the most informative relationships between the features of the model and the outcomes predicted[39]. The SHAP values assigned to each feature indicate the importance of the feature, with positive values suggesting a positive contribution to the prediction. In our study, many patients with a positive distribution of SHAP had greater age, male gender, and higher blood pressure, indicating them as the most important associated factors for hearing loss among CVD risk factors. The presence of well-known associated factors identified by our SHAP analysis strengthens the model's credibility for healthcare professionals evaluating its clinical applicability.

Various cross-sectional studies have demonstrated a link between hypertension and hearing loss, specifically in males but not females, which may be attributed to the protective effects of estrogen[17,47–51]. Changes in blood flow, tissue oxygenation, ionic balance in the cochlea, and microcirculatory insufficiency can collectively contribute to hearing loss in individuals with hypertension[50,52,53]. A recent meta-analysis has also highlighted the relationship between hearing loss and stroke, suggesting that previous cerebrovascular accidents (CVAs)

can impair blood flow to auditory structures[54]. CVD risk factors may cause hearing loss through damage to small auditory vessels from arteriosclerosis, reducing blood supply to structures like the cochlea[51,55]. The cochlea's high energy demands make it particularly vulnerable to conditions of reduced blood flow, which can impair its function[56]. Disruptions can directly damage hair cells, cause electrical issues, and reduce supporting cells, ultimately leading to hearing loss[55,57]. These pathophysiological links parallel those in CVDs like heart disease and stroke. Thus, cardiovascular conditions may reflect both systemic vascular problems and specific microvascular issues in the cochlea related to hearing[51].

The strengths of this study are manifold and bear significant clinical implications for medical practitioners across various specialties. By harnessing a large sample size from a nationally representative dataset, this research examined eight data-driven-based ML algorithms, including a uniquely designed MLNN model. The data-driven models in our study leverage the large and comprehensive NHANES dataset to discover patterns, enhance diagnostic accuracy, and enable personalized treatment plans. This methodology not only facilitates predictive analytics for identifying high-risk groups and implementing preventative measures but also streamlines healthcare operations, advances clinical research, and optimizes population health management[58–60]. Early detection of auditory dysfunction in individuals with modifiable risk factors is key to reducing hearing loss burden. Understanding specific and aggregate associated factors linked to hearing loss can inform guidelines advocating timely hearing evaluations and interventions for at-risk patients. This aligns with the growing personalized medicine emphasis, underscoring holistic care's importance. Prospective studies are needed to clarify CVD risk status impact on long-term hearing, enhancing understanding of clinical implications.

This study had some limitations. First, there were challenges in accurately capturing the comorbidity status of CVD on the exact day of the audiological evaluations, due to the non-concurrent scheduling of appointments for hearing assessments and other health evaluations. The cross-sectional nature of our dataset limited our ability to determine the causal relationship between CVD risk factors and hearing outcome. In addition, CVD risk factor data were derived from the closest available clinical encounter, which introduces an inability to account for potential fluctuations in these measures that may occur on a day-to-day basis. Factors like cardiovascular disease history and medication use were based on participant self-report, which may be subject to recall bias or underreporting. Moreover, the analysis did not specifically adjust for or isolate the individual effects of medication use, as the definition of CVD risk stratification inherently encompassed treatment modalities. Also, the lack of concurrent audiometric testing and assessment of objective hearing status is a constraint of the cross-sectional NHANES design. Lastly, while the algorithm was developed using data from the NHANES, this dataset might not yield precise CVD risk estimates across all racial and ethnic groups, potentially limiting the generalizability of the findings, as the NHANES cohort is specific to the United States. Future iterations of our predictive models should explore incorporating binaural hearing data or developing separate models to identify asymmetric hearing loss patterns, in order to more comprehensively capture an individual's overall hearing status beyond reliance on better ear pure-tone averages. A prospective study with an appropriately calculated sample size would be an essential next step to validate the promising findings from this retrospective analysis and further establish the clinical utility of the developed predictive models, considering prospective validation with concurrent audiometric and cardiovascular disease assessments in not only hearing loss, but also tinnitus. Although the study employed a rigorous methodology by separating the test dataset prior to model training, ensuring no data leakage, the evaluation of model performance remains limited to the NHANES dataset. This internal testing provides an unbiased estimate of performance but does not substitute for external validation. The outcomes may still exhibit optimism due to their dependence on the same dataset for both model development and testing. Future studies should aim to validate these models using independent, external datasets to confirm their generalizability and robustness in broader populations and real-world clinical scenarios.

Conclusion

In summary, our research marks a significant step in utilizing ML, particularly the MLNN model, for accurate hearing outcome prediction from CVD risk factors. The study highlights that MLNN and LightGBM outperform other ML models in predicting hearing outcomes and identifies key risk factors focusing on the incidence. These findings pave the way for incorporating advanced predictive models into clinical settings, offering a promising direction for personalized healthcare and early intervention strategies.

Data availability

The datasets analyzed during the current study are available in the National Health and Nutrition Examination Survey (NHANES) repository, [https://www.cdc.gov/nchs/nhanes/index.htm]. The data and code used in this study are available upon reasonable request to the corresponding author. We are committed to ensuring the reproducibility of our findings and welcome the opportunity to share our resources with interested researchers.

References

1. Haile, L. M. et al. Hearing loss prevalence, years lived with disability, and hearing aid use in the United States From 1990 to 2019: findings from the global burden of disease study. *Ear Hear* **45**, 257–267. https://doi.org/10.1097/aud.0000000000001420 (2024).
2. <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>
3. Haile, L. M. et al. Hearing loss prevalence and years lived with disability, 1990–2019: Findings from the Global Burden of Disease Study 2019. *Lancet* **397**, 996–1009. https://doi.org/10.1016/s0140-6736(21)00516-x (2021).
4. McDaid, D., Park, A.-L. & Chadha, S. Estimating the global costs of hearing loss. *Int. J. Audiol.* **60**, 162–170 (2021).

5. Škerková, M., Kovalová, M. & Mrázková, E. High-frequency audiometry for early detection of hearing loss: A narrative review. *Int. J. Environ. Res. Public Health* https://doi.org/10.3390/ijerph18094702 (2021).

6. Shim, H. S. et al. Metabolic syndrome is associated with hearing disturbance. *Acta Oto-Laryngologica* **139**, 42–47 (2019).

7. Baiduc, R. R., Sun, J. W., Berry, C. M., Anderson, M. & Vance, E. A. Relationship of cardiovascular disease risk and hearing loss in a clinical population. *Sci. Rep.* **13**, 1642 (2023).

8. Ramage-Morin, P. L., Gilmour, H., Banks, R., Pineault, D. & Atrach, M. Hypertension associated with hearing health problems among Canadian adults aged 19 to 79 years. *Health Rep.* **32**, 14–26. https://doi.org/10.25318/82-003-x202101000002-eng (2021).

9. Yang, J.-R. et al. Body mass index, waist circumference, and risk of hearing loss: A meta-analysis and systematic review of observational study. *Environ. Health Prev. Med.* **25**, 25. https://doi.org/10.1186/s12199-020-00862-9 (2020).

10. Bainbridge, K. E., Hoffman, H. J. & Cowie, C. C. Diabetes and hearing impairment in the United States: Audiometric evidence from the national health and nutrition examination survey, 1999 to 2004. *Ann. Intern. Med.* **149**, 1–10 (2008).

11. Alateeq, M., Alnizari, O. & Hafiz, T. A. Measuring the effect of smoking on hearing and tinnitus among the adult population in the kingdom of Saudi Arabia. *Cureus* **15**, e39689. https://doi.org/10.7759/cureus.39689 (2023).

12. Curti, S. A. et al. Relationship of overall cardiovascular health and hearing loss in the Jackson heart study population. *Laryngoscope* **130**, 2879–2884 (2020).

13. Sun, Y.-S. et al. Components of metabolic syndrome as risk factors for hearing threshold shifts. *PloS one* **10**, e0134388 (2015).

14. Mick, P. T. et al. Associations between cardiovascular risk factors and audiometric hearing: Findings from the Canadian Longitudinal Study on Aging. *Ear Hear.* **44**, 1332–1343 (2023).

15. Papadopoulou, A. M., Papouliakos, S., Karkos, P. & Chaidas, K. The impact of cardiovascular risk factors on the incidence, severity, and prognosis of sudden sensorineural hearing loss (SSHL): A systematic review. *Cureus* **16**, e58377. https://doi.org/10.7759/cureus.58377 (2024).

16. Shargorodsky, J., Curhan, S. G., Eavey, R. & Curhan, G. C. A prospective study of cardiovascular risk factors and incident hearing loss in men. *Laryngoscope* **120**, 1887–1891. https://doi.org/10.1002/lary.21039 (2010).

17. Tan, H. E. et al. Associations between cardiovascular disease and its risk factors with hearing loss-A cross-sectional analysis. *Clin. Otolaryngol.* **43**, 172–181. https://doi.org/10.1111/coa.12936 (2018).

18. Wimalarathna, H. et al. Machine learning approaches used to analyze auditory evoked responses from the human auditory brainstem: A systematic review. *Comput. Methods Progr. Biomed.* **226**, 107118 (2022).

19. Chen, F., Cao, Z., Grais, E. M. & Zhao, F. Contributions and limitations of using machine learning to predict noise-induced hearing loss. *Int. Arch. Occup. Environ. Health* **94**, 1097–1111. https://doi.org/10.1007/s00420-020-01648-w (2021).

20. Zhang, K. et al. Machine learning-based prediction of survival prognosis in esophageal squamous cell carcinoma. *Sci. Rep.* **13**, 13532. https://doi.org/10.1038/s41598-023-40780-8 (2023).

21. Cherkassky, V. & Mulier, F. M. *Learning from data: Concepts, theory, and methods* (John Wiley & Sons, 2007).

22. Goff, D., Lloyd-Jones, D. & Bennett, G. Erratum: 2013 ACC/AHA guideline on the assessment of cardiovascular risk: A report of the american college of cardiology/american heart association task force on practice guidelines. *J. Am. Coll. Cardiol.* **63**, 20003 (2014).

23. D'Agostino, R. B. et al. General cardiovascular risk profile for use in primary care. *Circulation* **117**, 743–753. https://doi.org/10.1161/CIRCULATIONAHA.107.699579 (2008).

24. Gathman, T. J., Choi, J. S., Vasdev, R. M., Schoephoerster, J. A. & Adams, M. E. Machine learning prediction of objective hearing loss with demographics, clinical factors, and subjective hearing status. *Otolaryngol.-Head Neck Surg.* **169**, 504–513 (2023).

25. Little, R. J. & Rubin, D. B. *Statistical analysis with missing data* Vol. 793 (John Wiley & Sons, 2019).

26. Alam, S., Ayub, M. S., Arora, S. & Khan, M. A. An investigation of the imputation techniques for missing values in ordinal data enhancing clustering and classification analysis validity. *Decis. Anal. J.* **9**, 100341. https://doi.org/10.1016/j.dajour.2023.100341 (2023).

27. Afkanpour, M., Hosseinzadeh, E. & Tabesh, H. Identify the most appropriate imputation method for handling missing values in clinical structured datasets: A systematic review. *BMC Med. Res. Methodol.* **24**, 188 (2024).

28. Heymans, M. W. & Twisk, J. W. R. Handling missing data in clinical research. *J. Clin. Epidemiol.* **151**, 185–188. https://doi.org/10.1016/j.jclinepi.2022.08.016 (2022).

29. Rodríguez, P., Bautista, M. A., Gonzalez, J. & Escalera, S. Beyond one-hot encoding: Lower dimensional target embedding. *Image Vis. Comput.* **75**, 21–31 (2018).

30. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intel. Res.* **16**, 321–357 (2002).

31. Muthukrishnan, R. & Rohini, R. in *IEEE international conference on advances in computer applications (ICACA).* 18–20 (IEEE) (2016)

32. Shamshirband, S., Fathi, M., Dehzangi, A., Chronopoulos, A. T. & Alinejad-Rokny, H. A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues. *J. Biomed. Inform.* **113**, 103627 (2021).

33. Joloudari, J. H. et al. Coronary artery disease diagnosis; Ranking the significant features using a random trees model. *Int. J. Environ. Res. Pub. Health* **17**, 731 (2020).

34. Jalali, A. et al. Deep learning for improved risk prediction in surgical outcomes. *Sci. Rep.* **10**, 9289 (2020).

35. Ketkar, N. & Santana, E. Introduction to Deep Learning. In *Deep learning with Python* Vol. 1 (ed. Ketkar, N.) (Springer, 2017).

36. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

37. Abe, D. et al. A prehospital triage system to detect traumatic intracranial hemorrhage using machine learning algorithms. *JAMA Netw. Open* **5**, e2216393–e2216393 (2022).

38. Zhang, K. & Demner-Fushman, D. Automated classification of eligibility criteria in clinical trials to facilitate patient-trial matching for specific patient populations. *J. Am. Med. Inform. Assoc.* **24**, 781–787 (2017).

39. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems* **30** (2017).

40. Pan, X. A. et al. survival prediction model via interpretable machine learning for patients with oropharyngeal cancer following radiotherapy. *J. Cancer Res. Clin. Oncol.* **149**(10), 6813–6825 (2023).

41. Zhao, Y. et al. Machine learning models for the hearing impairment prediction in workers exposed to complex industrial noise: A pilot study. *Ear Hear.* **40**, 690–699. https://doi.org/10.1097/aud.0000000000000649 (2019).

42. Charih, F., Bromwich, M., Mark, A. E., Lefrançois, R. & Green, J. R. Data-driven audiogram classification for mobile audiometry. *Sci. Rep.* **10**, 3962. https://doi.org/10.1038/s41598-020-60898-3 (2020).

43. Ellis, G. M. & Souza, P. E. Using machine learning and the national health and nutrition examination survey to classify individuals with hearing loss. *Front. Digit. Health* **3**, 723533 (2021).

44. Ke, G. *et al.* Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* **30** (2017).

45. Soylemez, E. et al. Predicting noise-induced hearing loss with machine learning: The influence of tinnitus as a predictive factor. *J. Laryngol. Otol.* **138**, 1030–1035. https://doi.org/10.1017/S002221512400094X (2024).

46. Radhakrishnan, S., Nair, S. G. & Isaac, J. Multilayer perceptron neural network model development for mechanical ventilator parameters prediction by real time system learning. *Biomed. Signal Process. Control* **71**, 103170 (2022).

47. Agrawal, Y., Platz, E. A. & Niparko, J. K. Prevalence of hearing loss and differences by demographic characteristics among US adults: Data from the National Health and Nutrition Examination Survey, 1999–2004. *Arch. Intern. Med.* **168**, 1522–1530 (2008).

48. Gates, G. A., Cobb, J. L., D'Agostino, R. B. & Wolf, P. A. The relation of hearing in the elderly to the presence of cardiovascular disease and cardiovascular risk factors. *Arch. Otolaryngol.Head Neck Surg.* **119**, 156–161 (1993).
49. Cruickshanks, K. J. et al. Hearing impairment prevalence and associated risk factors in the Hispanic Community Health Study/ Study of Latinos. *JAMA Otolaryngol.-Head Neck Surg.* **141**, 641–648 (2015).
50. Yikawe, S. S. et al. Hearing loss among hypertensive patients. *Egypt. J. Otolaryngol.* **35**, 307–312 (2019).
51. Wattamwar, K. et al. Association of cardiovascular comorbidities with hearing loss in the older old. *JAMA Otolaryngol. Head Neck Surg.* **144**, 623–629 (2018).
52. Ramatsoma, H. & Patrick, S. M. Hypertension associated with hearing loss and tinnitus among hypertensive adults at a tertiary hospital in South Africa. *Front. Neurol.* **13**, 857600 (2022).
53. Babarinde, J. A., Adeyemo, A. A. & Adeoye, A. M. Hearing loss and hypertension: Exploring the linkage. *Egypt. J. Otolaryngol.* **37**, 98. https://doi.org/10.1186/s43163-021-00162-1 (2021).
54. Tan, C. J. W. et al. Association between hearing loss and cardiovascular disease: A meta-analysis. *Otolaryngol. Head Neck Surg.* **170**, 694–707 (2024).
55. Helzner, E. P. et al. Hearing sensitivity in older adults: associations with cardiovascular risk factors in the health, aging and body composition study. *J. Am. Geriatr. Soc.* **59**, 972–979 (2011).
56. Ciorba, A., Chicca, M., Bianchini, C., Aimoni, C. & Pastore, A. Sensorineural hearing loss and endothelial dysfunction due to oxidative stress: Is there a connection?. *J. Int. Adv. Otol.* **8**, 16 (2012).
57. Choi, J. Y. et al. Age-associated repression of type 1 inositol 1, 4, 5-triphosphate receptor impairs muscle regeneration. *Aging (Albany NY)* **8**, 2062 (2016).
58. Nabavi, A., Safari, F., Kashkooli, M., Nabavizadeh, S. S. & Vardanjani, H. M. Early prediction of cognitive impairment in adults aged 20 years and older using machine learning and biomarkers of heavy metal exposure. *Curr. Res. Toxicol.* **7**, 100198 (2024).
59. Austin, R. R., McLane, T. M., Pieczkiewicz, D. S., Adam, T. & Monsen, K. A. Advantages and disadvantages of using theory-based versus data-driven models with social and behavioral determinants of health data. *J. Am. Med. Inform. Assoc.* **30**, 1818–1825 (2023).
60. Chorev, M. et al. *In MEDINFO 2017: Precision Healthcare through Informatics 332–336* (IOS Press, 2017).

Author contributions

Study concept and design: AN; Acquisition of data: AN and FS; Statistical Analysis: AN, MK, and LAC, Analysis and interpretation of data: AN, TA, and FS; Drafting of the manuscript: AN, MK, and FS, Critical revision of the manuscript for important intellectual content: AN, LAC, TA, MAK, and FS; Study supervision: AN, LAC, TA, MAK, and MK. All individuals listed as (co)-authors have met the authorship criteria, and nobody who qualifies for authorship is omitted from the list. The final manuscript was corrected and approved by all authors.

Declarations

Competing interests

The authors declare no competing interests.

Ethics and consent to participate

This study utilized publicly available deidentified data from the National Health and Nutrition Examination Survey (NHANES) 2012–2018. As the data analysis was performed on anonymous secondary data, ethics approval and participant consent were not required. The data collection protocols and consent procedures for the NHANES 2012–2018 were approved by the National Center for Health Statistics Research Ethics Review Board (approval number: Protocol #2005–06). Written informed consent was obtained from all NHANES participants prior to enrolment. This study complies fully with the Declaration of Helsinki ethical principles for medical research involving human subjects.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-94253-1.

Correspondence and requests for materials should be addressed to F.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.