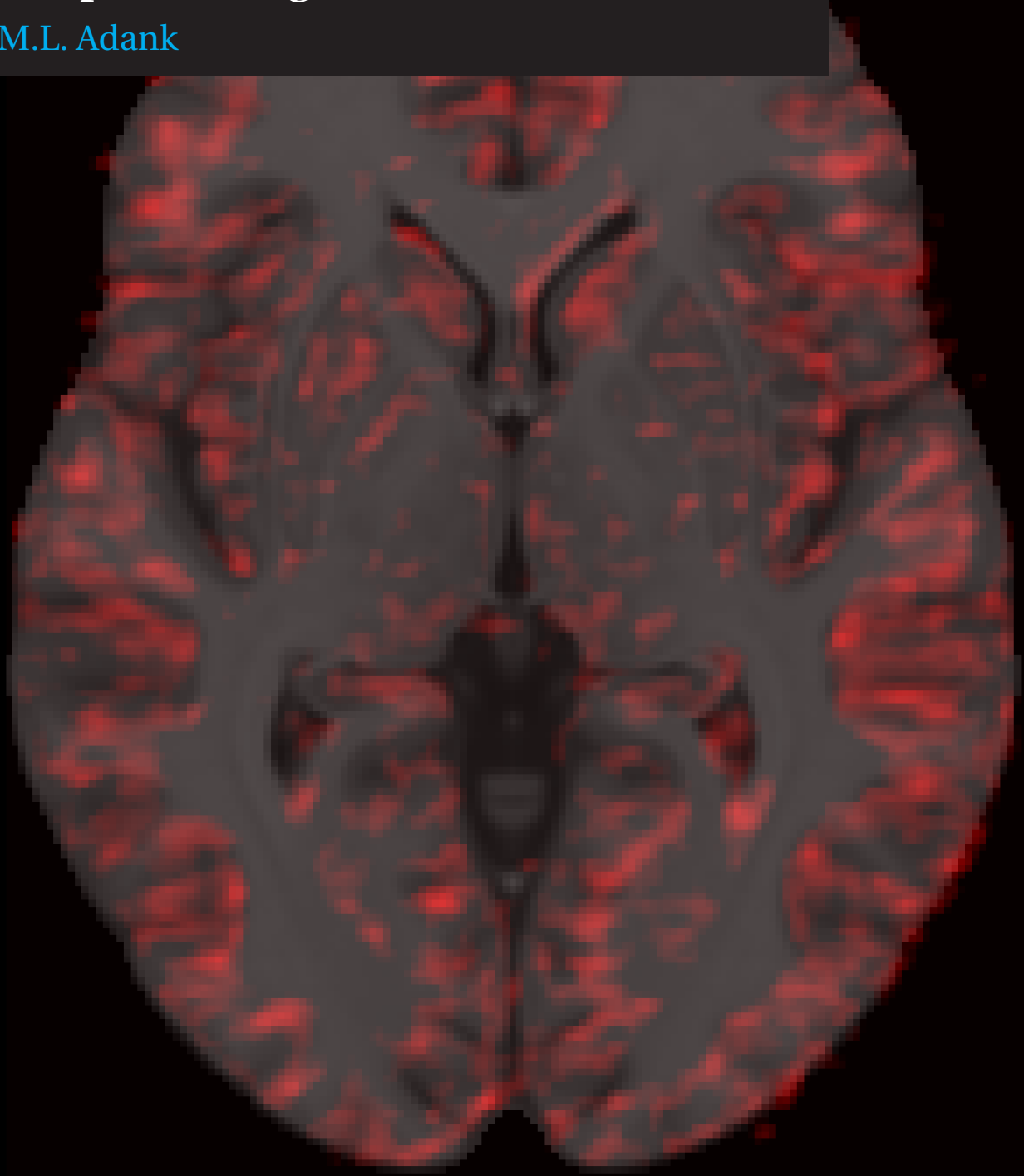
Analyzing gray matter differences in age-related hearing loss using multivariable linear regression and deep learning

M.L. Adank

Technische Universiteit Delft



TU Delft Delft University of Technology

Challenge the future

Analyzing gray matter differences in age-related hearing loss using multivariable linear regression and deep learning

by

M.L. Adank

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended on Friday December 20, 2019 at 3:00 PM.

Student number:	4222385	
Project duration:	February 1, 2019 – December 20, 2019	
Thesis committee:	Prof. dr. W. J. Niessen,	TU Delft & Erasmus MC
	Dr. F. M. Vos,	TU Delft
	Dr. V. Cheplygina,	TU Eindhoven
Supervisors:	Dr. E. E. Bron,	Erasmus MC
	P. H. Croll,	Erasmus MC
	A. Goedegebure,	Erasmus MC
	Prof. dr. M. W. Vernooij,	Erasmus MC
	Prof. dr. W. J. Niessen,	TU Delft & Erasmus MC

This thesis is confidential and cannot be made public until December 20, 2019.

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft

Preface

Before you lies the result of my graduation project. Herewith I complete my master's education in Biomedical Engineering, with Medical Physics as specialization. Almost eleven months of research and coding led to this report and the results presented in it. I look back on an intense and great experience in which I learned a lot.

I am grateful for the opportunity of doing research on the brain, a subject that already got my interest as a child. Gaining knowledge about deep learning and applying this technique to investigate brain differences in age-related hearing loss, was fascinating to me. I felt even more motivated to investigate this association, when seeing elderly people with hearing loss in the clinic, facing difficulties with communication in their daily lives.

First of all, I would like to thank Esther Bron and Pauline Croll for their great supervision during my thesis project. Esther, many thanks for your critical view on my work and your honest and realistic advice during my project. I highly appreciate your time and energy for supporting me during this journey. Pauline, many thanks for your help with writing and the fact that you were always available to discuss issues I was facing, even when you were having a very busy time for yourself as well.

Secondly, I would like to thank Wiro Niessen, André Goedegebure and Meike Vernooij for their time and valuable insights during my thesis project. Although our meetings were less regularly, your insights have been a great contribution to my work.

Thirdly, a word of thanks to Frans Vos and Veronika Cheplygina for taking on the task of reviewing my work as members of the examination committee. I truly appreciate the fact that I was allowed to hand-in my thesis a bit later due to the unfortunate illness in the last week of my project.

Finally, I would like to thank members from the Biomedical Imaging group for their help with my project and with coding. Without them, this would not have been such a joyful experience. I am also really grateful to my family, boyfriend, and friends who supported me in this journey and gave me distraction when needed. Thank you!

M.L. Adank
Rotterdam, December 2019

Abstract

Objective: Recent studies have suggested an association between age-related hearing loss and cognitive decline. Yet, the underlying mechanism explaining this relation remains unknown. In this regard, several studies investigated gray matter (GM) differences in age-related hearing loss but presented inconsistent results regarding the association and regions involved. To our knowledge, a data-driven approach for exploring this association has not been performed. Therefore, we aimed to investigate possible GM differences and regions involved in age-related hearing loss using conventional multivariable linear regression and deep learning.

Methods: Within the population-based Rotterdam Study, 2070 participants (mean age: 65.5 years) underwent pure-tone audiometry to quantify hearing thresholds (hearing loss [> 40 dB], n=205; normal-hearing controls [< 20 dB], n=822). Magnetic resonance (MR) imaging was performed to obtain GM volumes of the superior temporal and precentral gyrus, and GM modulated images. Using multivariable linear regression we investigated the associations between age-related hearing loss and GM volume in the superior temporal and precentral gyrus. A convolutional neural network (CNN) was trained to classify hearing loss and normal-hearing controls based on GM modulated images of the whole brain and the region around the superior temporal gyrus. Visualization of relevant features for the classification was performed with gradient-weighted activation mapping (Grad-CAM).

Results: We found that age-related hearing loss was significantly associated with smaller GM volumes in the right hemisphere of both the superior temporal gyrus (difference in standardized brain volume per dB increase: -0.006 [95% CI: -0.010, -0.003]) and precentral gyrus (difference: -0.005 [95% CI: -0.008, -0.001]). The CNN classification performance ranged between 0.89 and 0.96 area under the receiver-operating characteristic curves. Analysis of relevant features for the classification showed that features were not specific to the superior temporal gyrus or primary auditory cortex, but appeared across the whole brain. Furthermore, we noticed that misclassified subjects were significantly related to age.

Conclusion: This study shows that age-related hearing loss is related to both GM volume in the superior temporal and precentral gyrus. Moreover, relevant features for the classification of age-related hearing loss were observed across the whole brain. These results may be explained by a third factor affecting both hearing loss and neurodegeneration. As age likely is the third factor involved, a longitudinal study design or age-matched groups are required in further studies on age-related hearing loss.

Contents

1 Introduction

Age-related hearing loss, also known as presbycusis, is highly prevalent in older adults and affects approximately one-third of individuals over 65 years [77]. With a growing and aging population, the number of people with age-related hearing loss and its consequences will increase accordingly [23].

Age-related hearing loss is characterized by a loss of hearing sensitivity and reduced speech understanding in noisy environments [25]. In addition, the rate of auditory processing declines and individuals experience more difficulties with localizing sounds [25].

These symptoms appear due to damage in the peripheral or central auditory system (see Figure 1.1) [24, 25]. The peripheral system consists of the outer ear, middle ear, and the cochlea; the central auditory system involves the pathway from the cochlear neurons up to the primary auditory cortex. The primary auditory cortex is the brain region were auditory input is first perceived [32]. Damage in the peripheral system is seen in the cochlea, where the stria vascularis and the outer hair cells are mainly affected by age-related degeneration [25, 37]. Besides aging, environmental factors like occupational noise and smoking contribute to peripheral damage [24, 25, 64], complicating the causal-effect relations for the disease [23]. Other risk factors for age-related hearing loss include cardiovascular factors, ototoxic medication, and genetic susceptibility [25, 34, 45, 64].

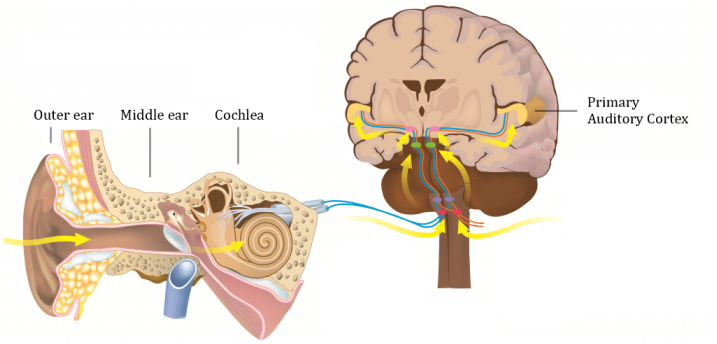


Figure 1.1: Peripheral (left) and central (right) auditory system [12]

As a result of the impoverished communication, age-related hearing loss may have a profound impact on a person's psychosocial and mental well-being and can lead to social isolation, loneliness, depression and a decreased quality of life [25, 44, 45, 56]. Recent epidemiological studies also identified an association between age-related hearing loss, cognitive decline and an increased risk of dementia [12, 23, 28, 31, 45, 51, 75]. However, the underlying mechanism between hearing loss, cognitive decline and dementia remains unknown.

Several hypotheses have been made to this regard (see Figure 1.2). First, brain changes or cognitive decline may induce hearing loss (see Figure 1.2a). Though little evidence for this 'cognitive load on perception' hypothesis exists from both animal and human structural magnetic resonance imaging (MRI) studies [39, 53, 75]. Second, age-related hearing loss may precede cognitive decline. Two hypotheses follow this potential relation. One is the 'information-degradation' hypothesis, which states that hearing loss leads to a degraded auditory signal and requires additional cognitive resources (see Figure 1.2b). These resources are then not available for other cognitive processes and may lead to a compromised cognition. Alternatively, the 'sensory-deprivation' hypothesis states that hearing loss leads to permanent neuroplastic changes and

cognitive decline due to poorer auditory input [51] (see Figure 1.2c). Finally, a 'common-cause' hypothesis is proposed, where a third-factor influences both age-related hearing loss and cognitive decline and may potentially lead to a false causal association between hearing loss and cognitive function [53, 60, 69] (see Figure 1.2d).

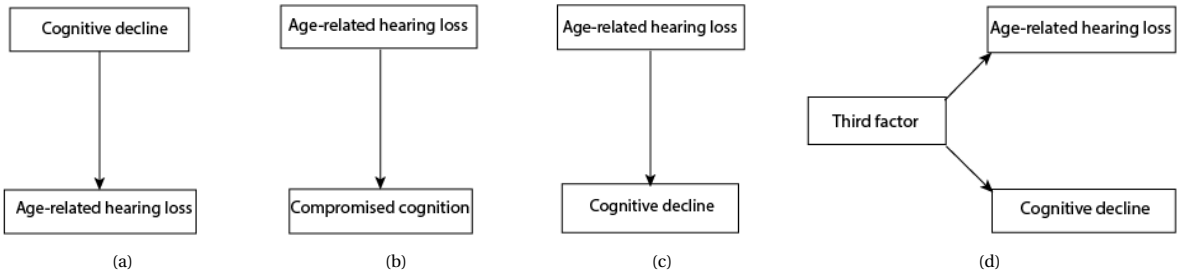


Figure 1.2: Summary of hypothesis explaining the relation between hearing loss and cognitive decline: A) Cognitive Load on Perception hypothesis, B) Information Degradation Hypothesis, C) Sensory Deprivation Hypothesis, D) Common cause hypothesis

As age-related hearing loss may induce brain changes and cognitive decline, the number of studies investigating structural brain differences in age-related hearing loss has increased. Some studies assessed whole brain and lobar differences in age-related hearing loss, and observed accelerated gray matter (GM) decline in the temporal lobe over time [50] or only white matter (WM) changes in all lobes [65]. Other studies investigated specific auditory and cognitive regions, and suggest an association between age-related hearing loss and GM differences in the primary auditory cortex [18, 50, 59, 62]. This association was mainly found in the right hemisphere [50, 59, 62]. Results regarding the association between hearing loss and other language and cognitive regions were inconsistent [4, 18, 50, 59, 62, 78]. Moreover, most studies had a small sample size and did not adjust for important confounding factors, like cardiovascular risk factors. A large population-based study is therefore needed to evaluate the observed differences in the auditory region while adjusting for important confounding factors. As the size and shape of the primary auditory cortex greatly differ among humans [6, 32, 33], the superior temporal gyrus, i.e. the region surrounding the primary auditory cortex, may serve as proxy for the auditory region.

Apart from hypothesis-driven research on brain differences in age-related hearing loss, it would be interesting to investigate possible brain regions involved in age-related hearing loss using a data-driven approach. So far, linear models were used to investigate a single imaging biomarker, like the GM volume or thickness, in age-related hearing loss. However, the complex interactions involved in age-related hearing loss may not be fully captured by a linear model and other imaging biomarkers, such as shape, could be predictive for the disease [1]. Deep learning is a multivariate method that can capture complex, non-linear patterns within the data and can assist in diagnosis and prediction of disease onset or progression [11, 54]. This data-analysis method has already proven to be successful for brain image analysis, like for the classification of Alzheimer's disease [54, 61]. In this task, the model learns important brain differences between MR-images of individuals with Alzheimer's and healthy controls to make an accurate prediction about the disease outcome. Despite good predictions, deep learning models are often difficult to interpret for humans [11, 54, 66]. Recent efforts have therefore been made to visualize features deemed relevant for the model in the classification task [72]. The purpose was to better comprehend the model and identify possible failure modes. However, visualization of the relevant features for disease prediction or diagnosis may as well give insight into regions that are possibly affected by a disease [9]. To our knowledge, no study has performed deep learning for age-related hearing loss.

In this study, we investigated the association between age-related hearing loss and GM volume in the superior temporal gyrus within the large population-based Rotterdam Study. To assess whether possible differences occur in regions that are independent of auditory function, we also examined the association between age-related hearing loss and GM volume in the precentral gyrus, or motor cortex.

Furthermore, we employed deep learning as a new analysis method to investigate possible brain differences and regions involved in age-related hearing loss. Hereto we performed a classification of age-related hearing loss based on MR-images of the whole brain and of the superior temporal gyrus. Relevant features for the classification model were visualized and missclassified images were examined.

2
Methods

2.1. Design and study population

This study is embedded within the Rotterdam Study, a large population-based cohort study in the Omnoord district in the city of Rotterdam, the Netherlands [35]. The Rotterdam Study was initiated in January 1990 and is aimed at investigating determinants of disease occurrence and progression in the elderly. From 2005 onwards, MRI has been incorporated into the study protocol and hearing assessments have been performed since 2011. This study comprises participants that underwent both MRI and hearing assessments between 2011 and 2014 (n=2878). From this group, we excluded participants from which no volume information of the superior temporal gyrus (n=458) or intracranial volume (ICV) (n=1) could be obtained due to incorrect MRI processing. We additionally excluded participants with asymmetric hearing loss (n=56), prevalent dementia (n=19), or participants for which the time between the MRI and hearing assessment was larger than 3 years (n=275). From the 2878 participants, 2070 participants were therefore eligible for multivariable linear regression analysis (See Figure 2.1).

For the classification of age-related hearing loss, we additionally excluded participants with mild hearing loss (n=957) and with an incorrect registration (n=86), leaving 1027 participants for analysis with deep learning (See Figure 2.1).

The Rotterdam Study was approved by the Medical Ethics Committee of the Erasmus MC and by the Ministry of Health, Welfare and Sport of the Netherlands. A written informed consent was obtained from all participants.

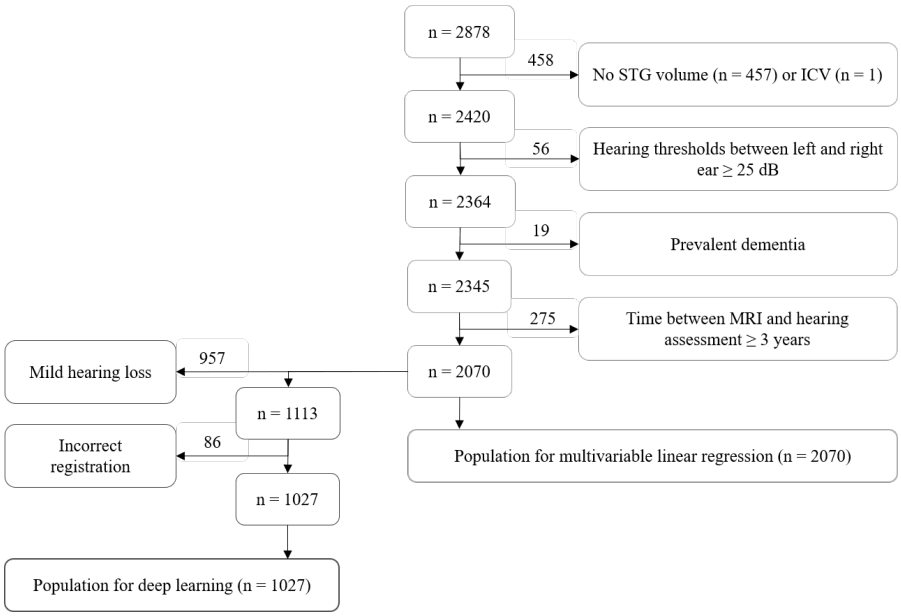


Figure 2.1: Flow chart of the included participants for analysis with multivariable linear regression and deep learning. Abbriviations: STG, superior temporal gyrus; ICV, intracranial volume; MRI, Magnetic Resonance Imaging.

2.2. Hearing assessment

Pure tone audiometry was performed to determine hearing thresholds of every participant. Hearing thresholds were obtained with a clinical audiometer (Decos audiology workstation, version 210.2.6, with AudioNigma interface) in a soundproof booth that met the International Organization for Standardization (ISO) standard 8253-1 [42]. Air conduction thresholds were determined at frequencies between 0.25 and 8 kHz. Bone conduction thresholds were obtained at 0.5 and 4kHz. All thresholds were measured using ISO-standard 8253-1 and masking was performed according to the method of Hood [36]. The best hearing ear was determined by taking the average air conduction over all frequencies. Low frequencies were determined as the average of 0.25, 0.5, and 1 kHz, high frequencies as the average of 2, 4, and 8 kHz. We categorized hearing loss in three categories according to the Bureau International d'Audiophonologie: normal hearing (0-20 dB), mild hearing loss (20-40 dB), moderate to severe hearing loss (>40 dB) [10]. Participants were examined on clinically relevant conductive hearing loss, which was considered to be present when the difference between air and bone conduction thresholds were larger than 15 dB. As age-related hearing loss is characterized by symmetric hearing loss [49], participants were also tested on asymmetric hearing loss. Asymmetry was here defined as a difference of 25 dB or more between hearing thresholds of the left and right ear.

2.3. Other measurements

Detailed information on relevant covariables was collected via physical examinations, interviews and blood sampling [34]. Glucose was determined enzymatically by the Hexokinase method. Systolic and diastolic blood pressure was determined using a random zero-sphygmomanometer. Total cholesterol and high-density lipoprotein cholesterol were measured from fasting blood samples [41]. Body mass index (BMI) was calculated by dividing weight (kg) by the square of height(m^2). Smoking was categorized as never, former, or current. Level of education was categorized by having completed primary education, lower, middle or higher education. Diabetes mellitus was stated present when the fasting blood glucose concentration was > 7.0 mmol/L and/or non-fasting blood glucose was > 11.0 mmol/L and/or when the participant used glucose-lowering medication. Hypertension was defined as having a systolic blood pressure >= 160 mmHg and a diastolic blood pressure of >= 90 mmHg or when using blood pressure-lowering medication. Hypercholesterolemia was stated present when the total cholesterol concentration was >= 6.2 mmol/L or when using lipid-lowering medication.

2.4. MRI acquisition

Within the Rotterdam Study, a multi-sequence MRI protocol is performed on a 1.5-Tesla MRI scanner (General Electric Healthcare, Milwaukee, USA). The protocol is extensively described before [40]. Briefly, it includes a T1-weighted (T1w), proton density-weighted (PDW), and fluid-attenuated inversion recovery (FLAIR) sequence for morphological imaging. The T1w image is acquired in 3D with full high in-plane resolution with thin slices (voxel size < 1 mm^3).

2.5. MRI preprocessing

2.5.1. Volume of the superior temporal gyrus

The GM volume of the superior temporal gyrus in the left and right hemispheres was obtained using the Freesurfer 5.1 pipeline. This pipeline involves preprocessing steps like motion correction and intensity normalization of the original T1-weighted images [71, 73]. Parcellation of the brain with the Freesurfer pipeline is performed using a surface-based model [21] and is described by the following steps. First, the pial GM surface is estimated from the surface of the WM segmentation and intensity information of the image. Using the information of the local curvature, a sulcal map is created, which then is registered to a spherical template [21]. At last, volume measures of anatomical regions of interest can be obtained from a spherical atlas that is mapped back to the individual sulcal map. Volumes of the superior temporal gyrus were acquired from the Desikan Killiany atlas, which is a gyral based surface atlas obtained from a data set of 40 MRI-scans [13]. Figure 2.2 shows the superior temporal gyrus of the right hemisphere obtained from Freesurfer with the Desikan Killiany atlas on the atlas template. The ICV per subject was quantified using an automated brain tissue classification method, based on a k-nearest-neighbor algorithm [74].

Figure 2.2: Desikan Killiany atlas with the superior temporal gyrus of the right hemisphere highlighted in green

2.5.2. GM modulated images

In principle, deep learning models have the ability to extract the relevant features from the image with minimal preprocessing of the image. However, previous research suggests that deep learning models may benefit from preprocessing steps like non-linear registration of the image [46]. Especially in the case of a small dataset, where it is more difficult for the model to train, the alignment of images to a common space may help the model to find the relevant features. GM modulated images were therefore used as input to our model. To obtain GM modulated images, the following steps were performed. First, the T1w-images were segmented into WM, GM and CSF using the k-nearest neighbor algorithm [74]. Second, the MR-images were non-linearly registered to the common GM probability template, ICBM MNI152GM template with a $1 \times 1 \times 1$ mm3 voxel resolution. To preserve information on absolute GM volume, a final spatial modulation step was performed where the image was multiplied with the Jacobian determinant of the transformation matrix [29]. The allowed range for the Jacobian determinant was set between 0.2 and 5 to avoid highly non-normal distributions that may occur when the Jacobian determinant has very small, or large values [55].

2.5.3. Correction for confounding factors

To classify individuals with hearing loss from normal hearing controls based on possible brain differences in MR-images, it is important to control for possible effects of confounding factors. Ideally, both classes are matched on confounding factors to ensure that they do not drive the classification. However, matched classes are not always possible, as is the case in the current study for age. GM volume shows a linear decline with increasing age [29]. Moreover, both sex and ICV are related to brain volume [8] and class differences in these factors could lead to confounding. Therefore we employed a voxel-wise linear regression to control for the confounding of age, sex, and ICV. This method was used as a preprocessing step and has previously been applied for age correction in dementia [17]. For each voxel, the regression coefficients β for the parameters age, sex and ICV, were determined based on the GM voxel intensities, y_C, of the control group. The linear regression model is described by the following equation, where the regression coefficients are estimated via the ordinary least squares (OLS) method; that is by minimizing the sum of squared residuals, $\sum \epsilon_c >$ min:

$$y_C = \beta_0 + \sum_{j=0}^{j=p} X_C \beta_j + \epsilon_C \tag{2.1}$$

β_0 is the intercept of the regression line, j accounts for the parameters age, sex and ICV ($p = 3$) and X_C is a matrix containing the parameter values and a constant for the control group. The contribution of each parameter was calculated by multiplying the regression coefficient with its deviation from the mean for each subject. Accordingly, the GM voxel intensities were adjusted by subtracting the contribution of each parameter from the original GM voxel intensity:

$$y_{cor} = y_{orig} - \sum_{j=0}^{j=p} \beta_j (X_j - \mu(X_{C,j})) \tag{2.2}$$

2.5.4. Image dimensions

After correction for confounding, images were cropped. For the auditory region, GM modulated images were cropped to a box that surrounds the superior temporal gyrus. The box was created by calculating the boundaries of the superior temporal gyrus for each subject, after which the largest boundaries were taken for the final coordinates of the box. In this way, the complete superior temporal gyrus for each subject was included. Deep learning analysis was performed on MR-images of the right superior temporal gyrus, which had a final box size of (60, 93, 87). For whole-brain analysis, the GM-modulated image (197, 233, 189) was cropped to exclude background pixels, which resulted in images with size (146, 182, 155).

2.6. Multivariable linear regression analysis

Multivariable linear regression was performed to investigate the association between age-related hearing loss (hearing threshold per dB increase) and standardized GM volume (Z-score GM volume) in the left and right hemispheres of the superior temporal gyrus. In addition, we evaluated the association between age-related hearing loss and standardized GM volume in the precentral gyrus. In the first model, we adjusted for age, age^2, sex, ICV and the time between MRI and auditory tests. To examine the effect of demographic and cardiovascular risk factors, we additionally adjusted for body mass index, alcohol intake, smoking status, educational level, mini-mental-state examination, diabetes mellitus, hypercholesterolemia and hypertension in the second model.

Missing data on hearing thresholds in 61 participants (0.03%) were imputed when no more than 2 hearing thresholds were missing per ear. The analyses were performed with IBM SPSS Statistics for Windows, version 24.0 (International Business Machines Corporation, Armonk, New York). The Bonferroni correction for multiple testing was applied to determine the statistical significance for each examined region (n=6), resulting in a significance level of $p < 0.0083$.

2.7. Deep learning analysis

Deep learning is a multivariable data-analysis method that transforms input (images) to output while learning higher-order features [11, 54]. In contrast to traditional machine learning algorithms that make use of 'handcrafted-features', deep learning models learn the relevant features from the data itself. In this study, a deep learning model was trained to classify individuals with age-related hearing loss and normal hearing controls based on brain differences (features) in MRI-scans. Accordingly, the learned features were visualized to identify possible brain differences and regions involved. The final step included an evaluation of misclassified images.

2.7.1. 3D CNN for the classification of age-related hearing loss

Convolutional neural networks (CNNs) are currently the best performing deep learning models for image classification [54]. These models apply convolution operations on images with a set of learnable kernels to generate feature maps. These kernels have a smaller spatial extent than the input image and are used at every position on the image, such that the number of learnable parameters is greatly reduced compared to traditional neural networks [30]. To classify age-related hearing loss based on 3-dimensional (3D) GM-modulated images, we therefore used a 3D CNN. The classification framework is schematically depicted in Figure 2.3, where the image is visualized as 2D for simplicity. Our model was based on a 3D CNN for the classification of Alzheimer's disease [46]. This model was optimized for the classification of age-related hearing loss (see Appendix B), which led to the final model that is presented in this section.

Network architecture The architecture of our 3D CNN consists of 4 alternatively stacked convolution and pooling blocks, followed by a convolution operation with kernel size 1x1x1, global average pooling, and a softmax layer. The convolution layer generates feature maps with kernels of size 3x3x3, where the number of feature maps is specified under the blocks in the image (see Figure 2.3). Zero padding was employed to control the spatial size of the output volume [47]. The feature maps were spatially down-sampled with the pooling operation, for which we applied a convolution operation with stride 2x2x2. To avoid overfitting and speed up convergence of the network, both convolution and pooling blocks contain dropout (rate = 0.5) and batch normalization [30, 43]. Non-linearity was induced by activation with the hyperbolic tangent at the end of each block. The convolution with kernel size 1x1x1 reduced the total number of feature maps to 2, after

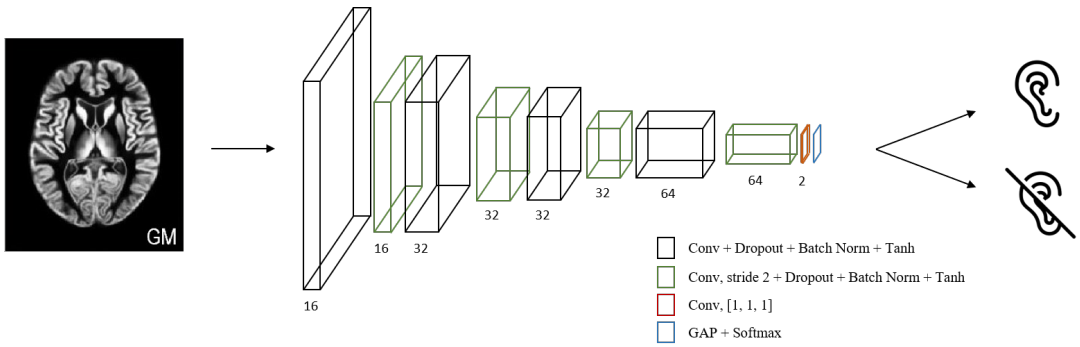


Figure 2.3: Framework for the classification of age-related hearing loss based on GM modulated images. The number underneath each block represent the number of filters in the convolution layer. Abbreviation: Conv, Convolution layer; Batch norm, batch normalization; tanh, hyperbolic tangent; GAP, global average pooling

which averaging by global average pooling further lowered the size of the feature maps to a single vector that could directly be fed into the final softmax layer. Using the softmax function, the final softmax layer generates a prediction score for both classes (age-related hearing loss or normal-hearing control).

Initialization The weights were randomly initialized using the Glorot initializer with a uniform distribution [26, 27]; biases were set to 0. The input images were standardized to have zero mean and unit variance. Due to memory constraints, the batch size for training the whole brain with image dimensions of (146, 182, 155) was equal to 4; the batch size for training the right superior temporal gyrus with image size (60, 93, 87) was 16.

Training The 3D CNN was trained and validated on 821 images (164 hearing loss, 657 controls) and tested on a separate dataset of 206 images (41 hearing loss, 165 hearing controls). During training, parameters are adjusted with the adaptive learning algorithm Adam to minimize the loss function [48]. Experiments revealed that the optimal initial learning rate was between 0.0005 and 0.001. The loss function quantifies how well the predicted class scores match the ground truth labels and is often given by the binary cross-entropy for a binary classification task [7, 47]. To account for class imbalance in our dataset, we used a variant on the binary cross-entropy, called the balanced cross-entropy loss, L_{BCE} [5, 52]:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N} \beta y_i log(p(y_i)) + (1-\beta)(1-y_i) log(1-p(y_i)) \tag{2.3}$$

where N is the number of samples in the batch over which the loss is calculated, β is the balancing term, y_i stands for the given class (0 in case of normal-hearing and 1 in case of hearing loss) and $p(y_i)$ is the predicted probability for that class. This loss function is equal to the binary cross-entropy when $\beta = 0.5$. However, for the classification of age-related hearing loss, β was optimal when $\beta = 1 - 164/657$, which is the inverse of the class frequency.

Evaluation To evaluate and optimize our model we used a 5-fold cross-validation with stratified shuffle split and the area under the receiver-operating-characteristic (ROC) curve (AUC) as performance metric on the test set. The stratified shuffle split method returns stratified random folds while preserving the percentage of samples for each class. The AUC is defined as the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance [20]. This performance metric has the advantage of being insensitive to changes in class distribution and is independent of the chosen classification threshold [20].

2.7.2. Visualization of relevant features with Grad-CAM

Gradient-weighted Class Activation Mapping, or Grad-CAM, was used to visualize the features deemed relevant for the model in the classification of age-related hearing loss. This visualization method uses the gradient of the class score, y^c, with respect to the feature maps, A^k, to provide visual explanations of the network. Grad-CAM is described by the following two formula's [72]:

$$\alpha_k^c = \frac{1}{Z}\sum_i\sum_j \frac{\delta y^c}{\delta A_{ij}^k} \tag{2.4}$$

$$L_{Grad-CAM}^c = ReLU(\sum_k \alpha_k^c A^k) \tag{2.5}$$

Equation 2.4 represents the importance of each feature map k for a specific class c and is given by the weights α. Global average pooling is performed to obtain an average weight per feature map, where Z stands for the number of voxels in the feature map. Equation 2.5 accordingly gives the localisation map per class, L^c, using a weighted combination of the weights and feature maps [72]. A $ReLU$ activation function is applied to highlight only positive features for a specific class. To highlight the discriminating features in the image, the localization map is scaled back to the original image size and shown as overlay on a reference brain. In the original method, Grad-CAM uses the gradient information and feature maps of the last convolutional layer [72]. However, the pooling operations result in smaller feature maps that do not provide location-specific information when scaled back to the original image. The third convolutional layer provided the best visual explanation of the network and was hence used for activation mapping.

2.7.3. Evaluation of missclassified subjects

The missclassified subjects were first evaluated using a visual inspection with Grad-CAM. This inspection yielded some speculations about the role that age could have played in the classification task. To that end, we performed the Welch's test to compare the mean age of the total correct and missclassified subjects in the two best performing models. The Welch's test is an alternative to the independent t-test when the groups have unequal variances and is given by the following formula [67, 76]:

$$t = \frac{\mu_1 - \mu_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \tag{2.6}$$

where μ_1 and μ_2 represent the mean age, s_1 and s_2 the variance of age, and N_1 and N_2 the amount of correct and missclassified subjects, respectively. The degrees of freedom, υ, can be approximated by the Welch-Satterthwaite equation:

$$\upsilon \approx \frac{(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2})^2}{\frac{s_1^4}{N_1^2(N_1-1)} + \frac{s_2^4}{N_2^2(N_2-1)}} \tag{2.7}$$

With the degrees of freedom and the T-table, we can find the minimum t-value that would correspond to a rejection of the null-hypothesis. Furthermore, Welch's test allows us to examine whether there is a statistically significant difference between the mean age of the correct and missclassified subjects.

2.7.4. Experiments

To gain confidence about the relevant features involved, our CNN was trained twice on both GM-modulated images of the whole-brain and of the superior temporal gyrus region. We selected the best fold of each training experiment and combined the Grad-CAM localization maps from all correctly classified subjects to obtain an average activation map per class for each experiment.

For whole-brain analysis, the classification was performed on images with and without additional GM extraction. This additional preprocessing step was executed to evaluate the influence of registration errors at the boundary of the image and within the WM region. GM extraction was employed on GM modulated images after correction for confounding and cropping.

Since no other study has performed machine learning for the classification of age-related hearing loss yet, a reference classification performance was lacking. Therefore, we employed a logistic regression in which

we classified individuals with age-related hearing loss based on gray matter volume in the right superior temporal gyrus. The logistic regression model uses the same model initialization and evaluation as deep learning and was used as baseline classification performance for analysis of the region around the superior temporal gyrus.

For both whole-brain analysis and analysis of the superior temporal gyrus region, we carried out a sanity check to verify whether the observed features depended on the training data. The sanity check corresponds to the 'data randomization test', where our CNN was trained on a copy of the dataset in which the labels were randomly permuted [2]. When activation mapping depends on the correct labeling of the classes, we should observe a difference in activation mapping of shuffled and correctly labeled data [2]. The randomly permuted labels yielded similar distributions of correct and false labels in both classes; the fraction of true and false labels in the hearing loss class was 0.190 and 0.202, respectively. For the normal-hearing control class, the fraction of true and false labels were 0.810 and 0.798.

2.7.5. Software

The network was trained in Keras with Tensorflow 1.12.0 as backend. All code for analysis with deep learning was written in Python 3.6.7.

3
Results

3.1. Multivariable linear regression analysis

Characteristics of the study population for analyzing GM differences in the superior temporal and precentral gyrus are shown in Table 3.1. The mean age of the population was 65.5 years, and 53.3% of the participants were women. The average audiogram of the best hearing ear of all participants is shown in Figure 3.1.

Table 3.1: Characteristics of the study population (N=2070)

Characteristics	
Age, years	65.5 ± 7.2
Sex, female, %	53.3
Best ear, left, %	53.1
Hearing threshold, dB	23.5 ± 12.0
ICV, mL	1137.9 ± 115.3
BMI, kg/m^2	27.3 ± 4.0
Alcohol intake, g/d	12.7 ± 15,6
Smoking status, %	
Never	32.0
Former	51.3
Current	16.3
Educational level, %	
Primary	6.8
Lower	35.5
Intermediate	30.8
Higher	26.5
MMSE test score	28.2 ± 1.7
Diabetes Mellitus, %	10.7
Hypertension, %	45.1
Hypercholesterolemia, %	54.9

dB, decibel; ICV, intracranial volume; BMI, Body mass index; g/d, grams/day; MMSE, mini-mental state examination: test to examine cognitive performance [0-30]

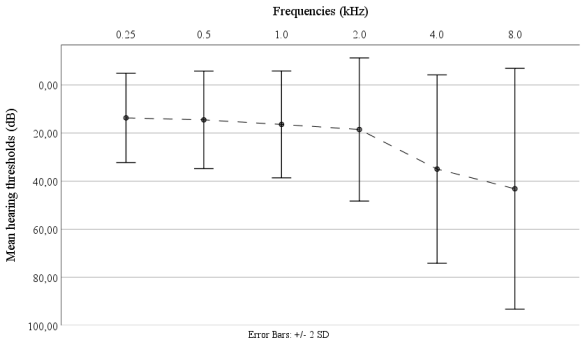


Figure 3.1: Average hearing thresholds of the best hearing ear (N = 2070)

In Table 3.2 we present the association between age-related hearing loss and GM volume in the superior temporal gyrus. We found that a higher hearing threshold, i.e. worse hearing function, was significantly associated with a lower GM volume in the superior temporal gyrus of the right hemisphere. We did not observe a significant association between age-related hearing loss and GM volume in the left hemisphere. The effect estimates for above associations were small and were strongest in the lower frequencies. Additional adjustment for cardiovascular risk factors did not change the effect estimates, although the associations in the lower and higher frequencies did not remain statistically significant after correction for multiple testing. Moreover, the difference in effect estimates between the left and right hemispheres in all frequencies became smaller after additional adjustment for cardiovascular risk factors.

Besides GM differences in the superior temporal gyrus, we observed a significant association between age-related hearing loss and GM volume in the right hemisphere of the precentral gyrus, i.e. our control region (see Table 3.3). Similar effect estimates were observed for the associations in both regions. Additional adjustment for cardiovascular risk factors did not change the effect estimate, though the association did not remain significant after multiple testing.

Table 3.2: Association of age-related hearing loss with gray matter volumes in the superior temporal gyrus using multivariable linear regression models (N=2070)

	Difference in gray matter volume (CI 95%) p-value, per dB increase	
	Left Hemisphere	Right hemisphere
All frequencies		
Model 1	-0.002 (-0.005, 0.001) 0.242	-0.006 (-0.010, -0.003) **0.000**
Model 2	-0.004 (-0.007, 0.000) 0.032	-0.005 (-0.009, -0.001) **0.007**
Low frequencies		
Model 1	-0.003 (-0.007, 0.001) 0.184	-0.007 (-0.011, -0.003) **0.001**
Model 2	-0.003 (-0.008, 0.002) 0.233	-0.006 (-0.011, -0.002) 0.009
High frequencies		
Model 1	-0.001 (-0.003, 0.001) 0.420	-0.004 (-0.006, -0.001) **0.004**
Model 2	-0.002 (-0.004, 0.001) 0.260	-0.003 (-0.006, 0.000) 0.024

CI, confidence interval; dB, decibel; GM, gray matter; WM, white matter
Model 1: adjusted for age, age^2, sex, intracranial volume, time between MRI and auditory tests. Model 2: Additionally adjusted for body mass index, alcohol intake, smoking status, educational level, mini-mental state examination, diabetes mellitus, hypercholesterolemia, and hypertension. Significant findings, $p < 0.0083$, are shown in bold.

Table 3.3: Association of age-related hearing loss with gray matter volumes in the precentral gyrus using multivariable linear regression models (N=2070)

	Difference in gray matter volume (CI 95%) p-value, per dB increase	
	Left Hemisphere	Right hemisphere
All frequencies		
Model 1	-0.004 (-0.008, 0.001) 0.009	-0.005 (-0.008, -0.001) **0.005**
Model 2	-0.004 (-0.008, 0.001) 0.022	-0.004 (-0.008, -0.001) 0.021

CI, confidence interval; dB, decibel; GM, gray matter; WM, white matter
Model 1: adjusted for age, age^2, sex, intracranial volume, time between MRI and auditory tests. Model 2: Additionally adjusted for body mass index, alcohol intake, smoking status, educational level, mini-mental state examination, diabetes mellitus, hypercholesterolemia, and hypertension.

3.2. Deep learning analysis

Characteristics of the hearing loss and normal-hearing control class are shown in Table 3.4. The average audiogram of both classes is presented in Figure 3.2. Besides the dissimilarity in the average hearing threshold between both classes, a large difference in mean age was observed; the mean age of the hearing loss class was 74.7 years, whereas the mean age of the normal hearing control class was 61.5 years. Furthermore, the percentage of women in the hearing loss class was lower (44.9 %) in comparison to the normal-hearing controls (61.2 %).

Table 3.4: Characteristics of the hearing loss and normal-hearing control class (N=1027)

Characteristics	Controls (N=822)	Hearing loss (N=205)
Age, years	61.5 ± 5.5	74.7 ± 8.0
Sex, female, %	61.2	44.9
Best ear, left, %	53.3	40.5
AVT, dB	13.3 ± 4.1	48.2 ± 8.3
ICV, mL	1130.1 ± 114.7	1140.2 ± 115.0

AVT, average hearing threshold; dB, decibel; ICV, intracranial volume

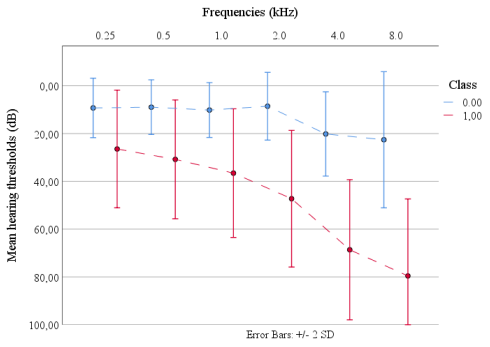


Figure 3.2: Average hearing thresholds of the best hearing ear for the hearing loss and normal-hearing control class (N = 2070)

3.2.1. Whole-brain classification

Figure 3.3 shows the classification performance for whole-brain analysis, with and without GM extraction, and its corresponding sanity check. For each experiment our CNN was trained twice, t1 and t2. The classification performance corresponds to the best performing model from the 5-fold cross-validation; individual model performance for all folds can be found in Appendix D. Repeated experiments resulted in a similar performance, which was achieved after less than 20 epochs training. A slightly lower classification performance was observed after GM extraction (see Appendix D), which also required a longer training time to reach convergence. The sanity check resulted in a model performance with an AUC ≈ 0.50 for both models.

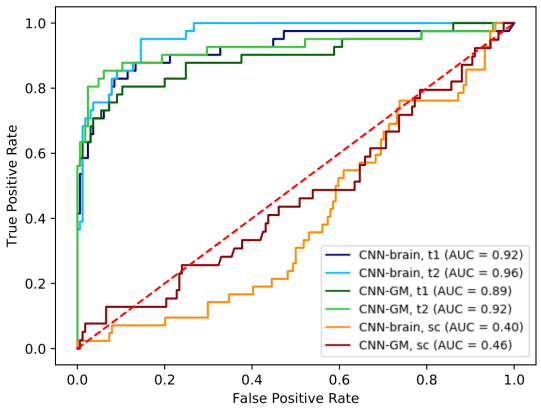


Figure 3.3: Receiver-operating-characteristic curves for the classification based on MR-images with and without GM extraction, CNN-brain and CNN-GM, and its corresponding sanity check. The CNN was trained twice with correct labels, t1 and t2. Abbreviation: sc, sanity check

3.2.2. Classification of the superior temporal gyrus region

Figure 3.4 shows the classification performance for analysis of the superior temporal gyrus region with our CNN and with logistic regression. The CNN was trained twice with correct labels, t1 and t2, and once with shuffled labels, i.e. the sanity check. Figure 3.4 shows the best performance of the 5-fold cross-validation; the individual performance across all folds can be found in Appendix D. Our CNN performed equally well in repeated experiments with correct labels and reached convergence within less than 10 epochs. Thereby, our model resulted in higher performance (AUC = 0.94) than classification with logistic regression solely based on GM volume in the superior temporal gyrus (AUC = 0.67). When our model was trained with shuffled labels, the classification performance was not better than guessing (AUC = 0.50).

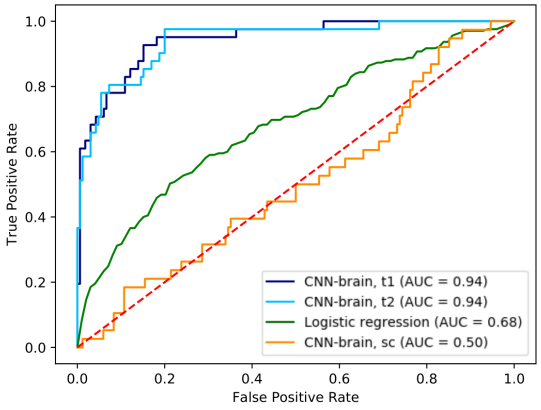
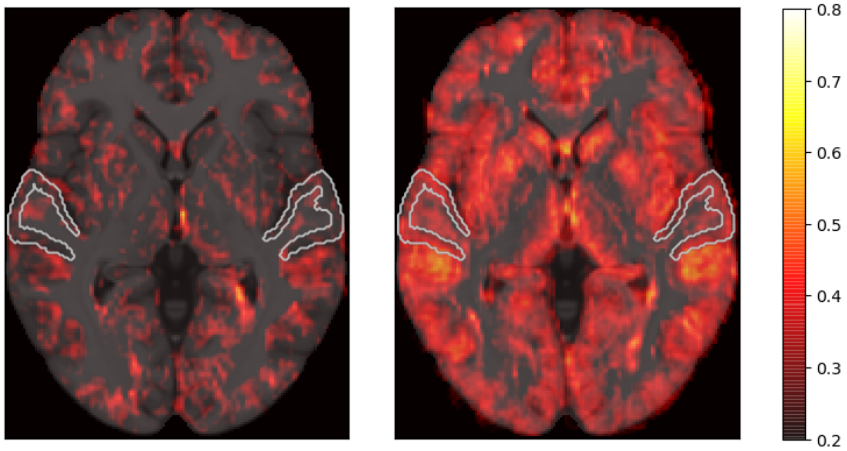


Figure 3.4: Receiver-operating-characteristic curves for the classification with our CNN and with logistic regression. The CNN was trained twice with correct labels, t1 and t2, and once with shuffled labels. Abbreviation: sc, sanity check.

3.2.3. Visualization of relevant features in the whole brain

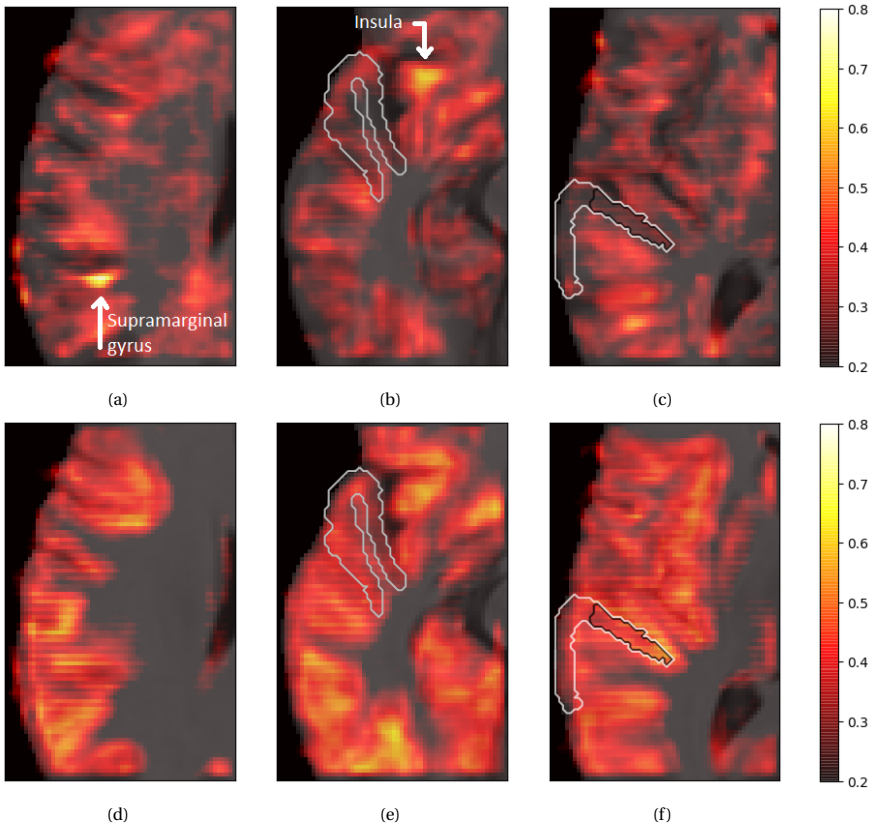
Grad-CAM visualizations for the classification of age-related hearing loss based on whole-brain images after GM extraction are presented in Figure 3.5. In both classes, activations were observed across the whole brain and were not specific to the superior temporal gyrus. Furthermore, the visualizations show lower activations in the hearing loss class in comparison to the control class. Along with these findings, we did not observe features specific to the superior temporal gyrus in the classification that was based on whole-images without GM extraction (see Appendix E.1). The sanity check resulted in different class activation maps E.4.1.



(a) (b)

Figure 3.5: Visualization of the average network activation in the whole brain based on MR-images with GM extraction, axial view; (a) activation of the hearing loss class, (b) activation of the normal-hearing control class. The white contour corresponds to the superior temporal gyrus, obtained from the Desikan Killiany atlas [13]. The color bar shows the normalized intensity values of the Grad-CAM image.

3.2.4. Visualization of relevant features in the superior temporal gyrus region



(a) (b) (c)

(d) (e) (f)

Figure 3.6: Visualization of the average network activation in the superior temporal gyrus region, axial view; a), (b) and (c) represent the hearing loss class at three positions; (d), (e) and (f) correspond to the activations of the hearing control class at the same three positions. The white matter contour corresponds to the superior temporal gyrus, obtained from the Desikan Killiany atlas [13]; the the black contour represents the primary auditory cortex, obtained from the Juelich atlas [19, 57]. The color bar shows the normalized intensity values of the Grad-CAM image.

Figure 3.6 shows the Grad-CAM visualizations for the average class activation in the superior temporal gyrus at three positions. Within the 3D activation map of the hearing loss class, two regions appeared to be of more importance. The first region is shown in (a) and relates to the supramarginal gyrus. In the normal-hearing control class (d), this region was less activated and did not stand out with respect to other brain regions. The second region where relevant features where observed, is visualized in (b) and involves the insula. Although similar activation values were found in the control class, the insular region again did not appear to be more relevant for the classification than other regions in the normal-hearing control class.

Figure 3.6 (c) and (f) visualizes the primary auditory cortex and the surrounding superior temporal gyrus. Despite the importance of these regions for auditory functioning, our Grad-CAM visualizations do not highlight these regions as relevant features for the classification of age-related hearing loss (b, c), (e, f).

In general, we observe less activation in the hearing loss class compared to the normal-hearing control class. Moreover, the visualizations show some activations in the WM region of the hearing loss class (a), which do not appear in the normal-hearing control class (d). The sanity check showed that the visualizations were data dependent (see Appendix E.4.2). Additional visualization results, including the corresponding sagittal and coronal view, from both whole-brain analysis and analysis of the superior temporal gyrus region can be found in Appendix E.2 and E.3.

3.2.5. Evaluation of misclassified subjects

We evaluated the subjects were the model gave a wrong prediction and observed significant differences in the mean age of the correctly classified and misclassified subjects (see Table 3.5). For both whole-brain analysis and analysis of the superior temporal gyrus region, subjects with hearing loss that were misclassified as hearing controls were younger in comparison to subjects that were correctly classified as hearing loss. The reverse effect was seen in normal-hearing controls, where misclassified subjects were older compared to the correctly classified subjects.

Furthermore, we noted that the mean age of misclassified subjects in the hearing loss class was lower than the mean age of the study population for this class (74.7 ± 8.0 years) and higher in case of the normal hearing controls (61.5 ± 5.5). For correctly classified subjects, the opposite relation holds, and smaller differences are found.

Table 3.5: Age differences between correct and misclassified subjects for each class based on MR-images of the whole brain and of the superior temporal gyrus region

	Hearing loss	Controls
Whole brain		
Correctly classified	79.1 ± 6.7 (n=56)*	60.8 ± 5.4 (n=321)*
Missclassified	68.3 ± 5.1 (n=26)*	70.1 ± 2.9 (n=9)*
Superior temporal gyrus region		
Correctly classified	76.7 ± 7.0 (n=77)*	59.2 ± 4.6 (n=279)*
Missclassified	64.8 ± 5.1 (n=5)*	68.1 ± 2.8 (n=51)*

*: significant difference, $p < 0.01$, in mean age between correct and missclassified subjects of each class

4
Discussion and conclusion

In this study, we analyzed GM differences in age-related hearing loss using both multivariable linear regression and deep learning. We found that higher hearing thresholds were associated with lower GM volume in both the superior temporal and the precentral gyrus, while the precentral gyrus is not implicated in auditory function. Using deep learning, we found that relevant features for the classification of age-related hearing loss were not specific to the superior temporal gyrus, but appeared across the whole brain. For the region around the superior temporal gyrus, our Grad-CAM visualizations suggest GM differences within the insula and supramarginal gyrus, and not in the primary auditory cortex. Interestingly, we noticed that misclassified subjects were strongly related to age.

4.1. Multivariable linear regression

We found that age-related hearing loss was significantly associated with decreased GM volume in the right hemisphere, but not in the left hemisphere. This observed hemispheric asymmetry is in agreement with previous studies [50, 59, 62]. Yet, some studies indicate GM differences in the auditory cortex of both hemispheres [18], or no GM differences at all [4, 78]. The latter two studies did not measure the hearing function of the control group and hence could have underestimated group differences. The right hemisphere also appears to be more affected by aging than the left hemisphere [63]. As most studies were of a cross-sectional study design, residual confounding by age could have been present. One longitudinal study reported accelerated GM volume decline in the right temporal lobe but did not correct for quadratic effects of age. Thus, it remains unclear whether the observed associations in the right hemisphere can be attributed to hearing loss or age-related neurodegeneration.

From these studies, just one study investigated the association between age-related hearing loss and GM volume in a control region [59]. The study did not observe an association between hearing loss and GM volume in the motor cortex. We examined the same control region and observed a significant association between age-related hearing loss and GM volume in the precentral gyrus. The discrepancy between these findings may be explained by the fact that we studied the superior temporal gyrus as auditory region and calculated Z-scores to compare regions of different sizes, whereas the other study compared differences in absolute volume of the primary auditory cortex with the absolute volume of the motor cortex. Given the small spatial extent of the primary auditory cortex, differences may appear faster with respect to the larger motor cortex. Especially with a sample size of 25, it is not surprising that small differences in the motor cortex were not observed.

Finally, we noticed stronger effect estimates for lower frequency thresholds compared to the higher frequency thresholds; differences in standardized brain volume per dB increase were -0.006 for lower and -0.003 for higher frequencies. Different types of age-related hearing loss have different origins and result in diverse audiograms [25, 70]. Sensory hearing loss is caused by loss of the outer hair cells and shows a decline in the higher frequencies. Metabolic hearing loss is caused by degeneration of the stria vascularis and affects the lower frequencies. The stronger association in the lower frequencies may be therefore be related to metabolic hearing loss. In that regard, various types of age-related hearing loss may lead to distinct brain changes (see Appendix C). Future research on this topic is needed to investigate potential differing relations between distinct forms of hearing loss with brain volume.

4.2. Deep learning

To our knowledge, we are the first study to use deep learning for the classification of age-related hearing loss. For both whole-brain analysis, and for the analysis of the superior temporal gyrus, we were able to achieve performances of 0.96 and 0.94 AUC, respectively. Our CNN thereby outperformed the classification

with a logistic regression model that was based solely on GM volume in the superior temporal gyrus and had an AUC of 0.68. Given a relatively shallow network and a classification performance of more than 0.90 after a few epochs, one could argue whether a complex deep learning model was necessary for this problem. Traditional machine learning models, like a support vector machine or even logistic regression, may perform equally well when presented with a set of relevant features. However, in this study we aimed to investigate GM differences without prior knowledge about the relevant features or regions of interest. Our CNN was able to extract relevant features from the data itself and revealed that discriminating features for the classification of age-related hearing loss were not specific to the superior temporal gyrus, but appeared along the whole brain.

Furthermore, our model generated new hypotheses about possible differences in the supramarginal gyrus and the insula. The supramarginal gyrus is involved in language processes [3]. Even though this region is not mentioned by previous studies, it could be an interesting region to investigate in more depth. The insula is known to be affected by aging [29, 63], and activation in this region suggests that effects of age may still be present in the GM modulated images after correction of confounding with the voxel-based linear regression. This hypothesis is supported by the characteristics of the misclassified subjects, who appear to be strongly age-dependent.

Interestingly, the Grad-CAM visualizations of the superior temporal gyrus also showed activations in the WM region of the hearing loss class. In principle, GM modulated images contain intensity values of zero in this region. However, after investigating this region in more depth, we indeed observed non-zero pixel values in subjects with hearing loss. This may be explained by the registration to the standard ICBM MNI152GM template. This template is based on 152 subjects with an age range of 18-44 years [22], whereas our study population has a mean age of 65.5 years. Registration to this template could lead to registration errors in elderly people with age-related neurodegeneration. As the mean age of subjects in the hearing loss class is larger (74.7 years) compared to the controls (61.5 years), registration errors were likely noticed as feature for the hearing loss class, but correspond to morphological brain changes caused by aging.

Besides the effects of age, non-specific activation patterns in the superior temporal gyrus or primary auditory cortex may be explained by another third factor. Vascular mechanisms can influence both hearing thresholds and brain health and could be the third factor involved. This idea is supported by the results from our multivariable linear regression, which implicated a possible role of the metabolic phenotype of age-related hearing loss. Furthermore, we observed an increase in classification performance, from 0.81 to 0.89 AUC, after correction for the confounding of age with the voxel-based linear regression (see Appendix B). When the classification of age-related hearing loss was solely driven by age, an increase in performance after age correction is not expected, and hence may be explained by another third factor.

4.3. Limitations and future work

Some limitations regarding the study population should be acknowledged. We assigned all participants with hearing loss in our population (> 45 years) with age-related hearing loss, but some participants may have developed hearing loss early in life. As the prevalence of hearing loss for individuals younger than 40 years old comprises less than 2 percent [58], the number of individuals incorrectly assigned with age-related hearing loss can be assumed very small. Another limitation regarding the study population is that participants with mild hearing loss are excluded in our deep learning analysis. Mild hearing loss is the most common form of hearing loss in the elderly [58] and by excluding this group, we could have overestimated GM differences. Yet, the results from our deep learning analysis are in line with the regression analysis, which included participants with mild hearing loss. Both analyses revealed that GM differences were not specific to the superior temporal gyrus and are likely influenced by a third factor.

Improvements can be made with respect to the preprocessing steps. First, the observed registration errors in our study could be avoided by using a study-specific template. Second, GM density maps are preferable for the segmentation of GM, which was based on a binary segmentation in this study. At last, a multivariable correction for confounding of age would be beneficial for our multivariable data analysis with deep learning. A recent study proposed a spatio-temporal reference model of the aging brain to assess morphological differences due to aging [38]. Although not designed for the correction of confounding, a correction based on a reference deformation field for aging may better capture morphological differences than a linear correction solely based on volume. However, age-matched classes are still preferable over correction methods and are recommended in future research on GM differences in age-related hearing loss.

Visualization of the features used in a neural network is still an important challenge in deep learning. Even

though methods like Grad-CAM provide insight into the location of relevant features, the features themselves remain unknown [66]. Hence, a complete description of the important aspects during classification cannot be given. Unlike some other visualization methods, Grad-CAM was dependent on both the model parameters and the data the model was trained on [2]. However, the scaling of the feature maps to the original image, makes Grad-CAM less location-specific. GP-Unet is a weakly supervised detection algorithm that can generate activation maps with full input-resolution by means of an encoder-decoder architecture [15, 16]. Although it requires a change in network architecture, GP-Unet could be interesting to use in future work as it has the potential to provide a more precise activation map.

In future work, it would also be desirable to have an MR-scanner with a higher field strength. The 1.5 Tesla scanner used in this study is common for both research and clinical practice but is less sensitive to smaller brain structures that may be present in age-related hearing loss. A higher field strength is mainly advantageous in future research with age-matched groups, where possible small GM differences related to hearing loss are not overshadowed by age-related neurodegeneration. In that case, our CNN may require more convolution layers but could prove helpful to detect small brain differences in age-related hearing loss.

Finally, a longitudinal population-based study is warranted to investigate intra-individual brain changes differences and elucidate longitudinal associations between age-related hearing loss and potential neurodegeneration independent of age.

4.4. Conclusion

In this study, we analyzed GM differences in age-related hearing loss using conventional multivariable linear regression and deep learning. Results from both methods revealed that GM differences were not specific to the superior temporal gyrus, but were present in other brain regions as well. Furthermore, we observed that misclassified subjects were significantly related to age and that age-related features were relevant for the classification. Although a direct effect of age-related hearing loss on neurodegeneration may be present, our findings suggest that these differences are negligible with respect to a third factor affecting the whole brain. As age likely is the third factor involved, a longitudinal study design or age-matched groups are required in further studies on age-related hearing loss.

A

Correction for confounding

To evaluate the voxel-wise correction for confounding, we visualized the β coefficients of the multivariable linear regression for the confounding of age (see Figure A.1). A negative β value corresponds to a decrease in volume with age; a positive β value relates to an increase in volume with age.

Although some voxels showed small positive β values, in general, we observed a small decrease in volume with age. These results are in line with research on age-related neurodegeneration, which shows a linear GM decline with aging [29, 63].

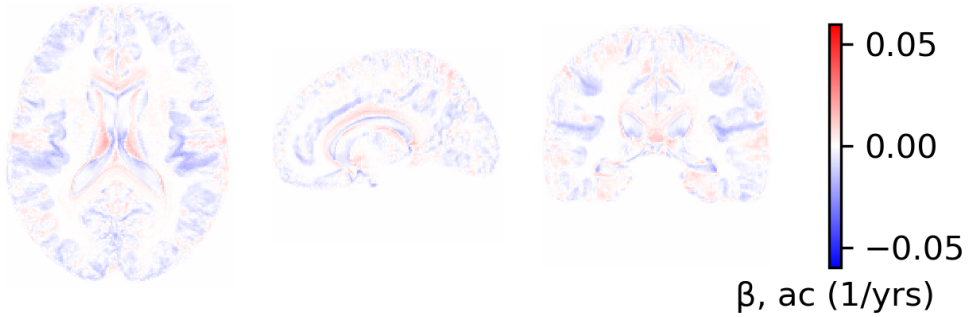


Figure A.1: β coefficients for the correction of age. Abbreviation: ac, age correction; yrs, years.

B
Model Optimization

This chapter shows the experiments that contributed most to the optimization of our model or were important for our classification task.

B.1. Preprocessing: correction for confounding

As described in section 2.5.3, the correction for confounding of age, sex and ICV was necessary for the classification of age-related hearing loss. This preprocessing step, with the correction of age in particular, also yielded a higher network performance (see Table B.1). As age may drive the classification before correction of confounding, the gain in performance could be explained by older subjects without hearing loss that are now correctly classified as hearing controls. Similarly, younger subjects with hearing loss would be classified as hearing controls when the classification is driven by age, but are correctly classified as hearing loss after correction of confounding.

Table B.1: Network performance with and without correction of confounders

Correction for confounding	AUC	# Filters in each conv block	Activation function	Loss function	Regularization
-	0.81 ± 0.03	16-32-32-64-2	tanh	BCE, $\beta \sim 0.75$	D + BN
age	0.89 ± 0.03	16-32-32-64-2	tanh	BCE, $\beta \sim 0.75$	D + BN
age, sex, ICV	**0.92 ± 0.02**	16-32-32-64-2	tanh	BCE, $\beta \sim 0.75$	D + BN

Abbreviations: AUC, area under the receiver-operating-characteristic curve, tanh, hyperbolic tangent; BCE, balanced cross entropy; D, dropout; BN, batch normalization.

B.2. Model Architecture

Starting with the network architecture proposed for the classification of Alzheimer's disease, we set up several experiments in which we varied the number of convolution layers and filters involved (see Table B.2). The CNN with 4 blocks of convolution and pooling with filter sizes 16, 32, 32 and 64 and a final convolution block with 2 filters achieved the best performance (AUC = 0.93) and was hence used as our model.

Table B.2: Network performance for varying model depths and number of filters

Correction for confounding	AUC	# Filters in each conv block	Activation function	Loss function	Regularization
age, sex, ICV	0.90 ± 0.02	8-16-32-32-2	tanh	BCE, $\beta \sim 0.75$	D + BN
age, sex, ICV	**0.93 ± 0.03**	**16-32-32-64-2**	tanh	BCE, $\beta \sim 0.75$	D + BN
age, sex, ICV	0.91 ± 0.02	16-32-64-64-2	tanh	BCE, $\beta \sim 0.75$	D + BN
age, sex, ICV	0.90 ± 0.01	16-32-32-64-64-2	tanh	BCE, $\beta \sim 0.75$	D + BN
age, sex, ICV	0.91 ± 0.01	16-32-32-64-64-32-2	tanh	BCE, $\beta \sim 0.75$	D + BN

The conv block consist of both the convolution and pooling block, with the exception of the last conv layer that only has a convolution operation. Abbreviations: AUC, area under the receiver-operating-characteristic curve; conv, convolution; tanh, hyperbolic tangent; BCE, balanced cross entropy; D, dropout; BN, batch normalization.

B.3. Activation function

The rectified linear unit (ReLU) is the most common activation function in CNNs [54]. Although similar classification performances could be obtained with a ReLU activation, the hyperbolic tangent (tanh) appeared more robust and was therefore used as activation function in our model (see Table B.3).

Table B.3: Network performance for activation with the hyperbolic tangent or ReLU

Correction for confounding	AUC	# Filters in each conv block	Activation function	Loss function	Regularization
age, sex, ICV	0.92 ± 0.02	16-32-32-64-2	tanh	BCE, $\beta \sim 0.75$	D + BN
age, sex, ICV	0.88 ± 0.06	16-32-32-64-2	ReLU	BCE, $\beta \sim 0.75$	D + BN
age, sex, ICV	0.93 ± 0.05	16-32-64-64-2	ReLU	BCE, $\beta \sim 0.75$	D + BN
age, sex, ICV	0.89 ± 0.05	16-32-64-64-2	ReLU	BCE, $\beta \sim 0.75$	D + BN

Abbreviations: AUC, area under the receiver-operating-characteristic curve; conv, convolution; tanh, hyperbolic tangent; BCE, balanced cross entropy; D, dropout; BN, batch normalization.

B.4. Loss function

As described in section 2.7.1, we used the balanced cross entropy as loss function to account for class imbalance in our dataset. A grid search revealed that the hyperparameter β = was optimal when $\beta \sim 0.75$ (AUC = 0.93), which equals the inverse class frequency (see Table B.4).

Table B.4: Network performance for different values of the balancing term, β, in the balanced cross entropy loss function

Correction for confounding	AUC	# Filters in each conv block	Activation function	Loss function	Regularization
age, sex, ICV	0.90 ± 0.02	16-32-32-64-2	tanh	BCE, β =0.50	D + BN
age, sex, ICV	0.89 ± 0.03	16-32-32-64-2	tanh	BCE, β =0.66	D + BN
age, sex, ICV	**0.93 ± 0.02**	16-32-32-64-2	tanh	**BCE, $\beta \sim 0.75$**	D + BN
age, sex, ICV	0.91 ± 0.02	16-32-32-64-2	tanh	BCE, β =0.90	D + BN

Abbreviations: AUC, area under the receiver-operating-characteristic curve; conv, convolution; tanh, hyperbolic tangent; BCE, balanced cross entropy; D, dropout; BN, batch normalization.

B.5. Regularization

Finally, we observed that the inclusion of batch normalization and dropout improved network performance of our CNN (see Table B.5). These regularization methods were therefore both integrated in our model architecture.

Table B.5: Network performance with and without regularization of dropout and batch normalization

Correction for confounding	AUC	# Filters in each conv block	Activation function	Loss function	Regularization
age, sex, ICV	0.89 ± 0.02	16-32-32-64-2	tanh	BCE, $\beta \sim 0.75$	-
age, sex, ICV	0.57 ± 0.05	16-32-32-64-2	tanh	BCE, $\beta \sim 0.75$	D
age, sex, ICV	0.91 ± 0.04	16-32-32-64-2	tanh	BCE $\beta \sim 0.75$	BN
age, sex, ICV	**0.93 ± 0.02**	16-32-32-64-2	tanh	BCE, $\beta \sim 0.75$	**D + BN**

Abbreviations: AUC, area under the receiver-operating-characteristic curve; conv, convolution; tanh, hyperbolic tangent; BCE, balanced cross entropy; D, dropout; BN, batch normalization.

C

Phenotypes age-related hearing loss

To investigate the hypothesis of differing relations between distinct forms of hearing loss with brain volume, we analyzed the association between metabolic and sensory hearing loss (average hearing threshold per dB increase) and standardized GM volume (Z-score GM volume) in the superior temporal gyrus.

Figure C.1 shows the typical audiograms per phenotype. To our knowledge, no clear definition of the hearing thresholds per phenotype exists. Therefore, the typical audiograms per phenotype were used to classify participants into distinct phenotypes. Metabolic hearing loss was stated present when the difference between low and high-frequency hearing thresholds was lower than 15 dB (n = 1009). Participants were assigned with sensory hearing loss when the low-frequency hearing thresholds were lower than 20 dB and the high-frequency hearing thresholds were larger than 40 dB (n = 385).

Table C.1 shows the association between metabolic and sensory hearing loss and GM volume in the superior temporal gyrus. We found that a higher hearing threshold was significantly associated with a lower GM volume in the right hemisphere of the superior temporal gyrus in both phenotypes. However, in the metabolic phenotype, we observed larger effect estimates in the right hemisphere compared to the left hemisphere, whereas the inverse was true for the sensory phenotype. These results suggest possible differing relations between distinct forms of hearing loss with brain volume.

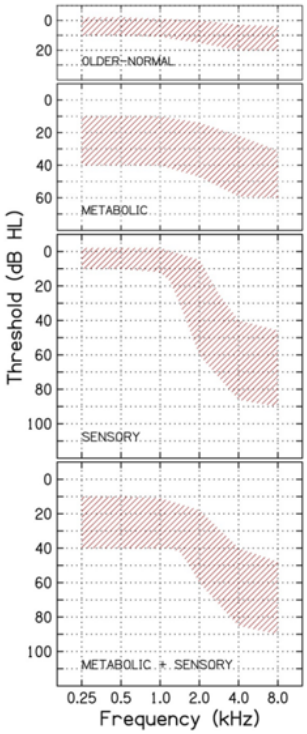


Figure C.1: Phenotypes for age-related hearing loss [14, 68]

Table C.1: Association of metabolic and sensory hearing loss with gray matter volumes in the superior temporal gyrus using multivariable linear regression (N=2070)

	Difference in gray matter volume (CI 95%) p-value, per dB increase	
	Left Hemisphere	Right hemisphere
Metabolic (n=1009)		
Model 1	-0.005 (-0.011, 0.001) 0.099	-0.011 (-0.018, -0.005) **0.001**
Model 2	-0.006 (-0.014, 0.001) 0.082	-0.010 (-0.018, -0.003) **0.008**
Sensory (n=385)		
Model 1	-0.011 (-0.027, 0.005) 0.188	-0.007 (-0.022, 0.009) 0.393
Model 2	-0.020 (-0.039, -0.001) 0.036	-0.016 (-0.034, 0.002) **0.008**

Abbreviations: CI, confidence interval; GM, gray matter; WM, white matter; Hearing thresholds were taken over all frequencies.
Model 1: adjusted for age, age^2, sex, intracranial volume, time between MRI and auditory tests. Model 2: Additionally adjusted for body mass index, alcohol intake, smoking status, educational level, mini-mental state examination, diabetes mellitus, hypercholesterolemia and hypertension. Significant findings, $p < 0.0083$, are shown in bold.

D Additional results of the classification performance

D.1. Classification of the superior temporal gyrus

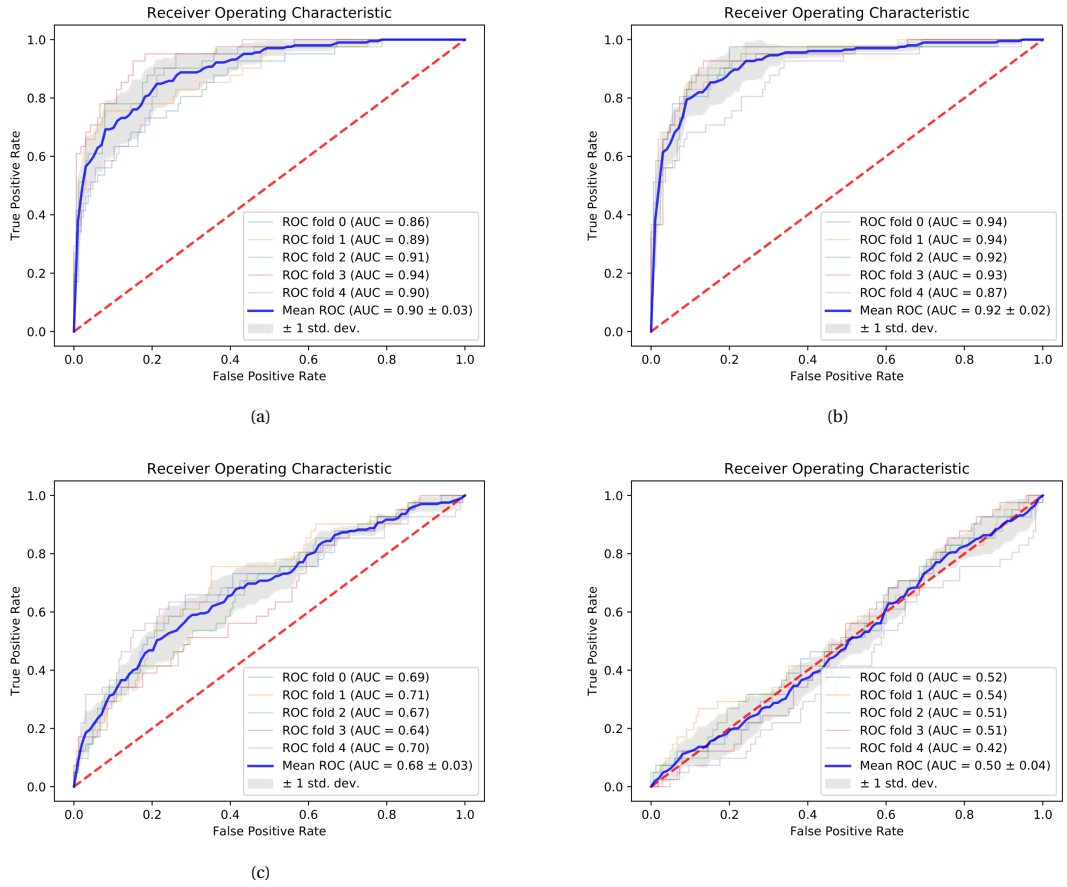


Figure D.1: Receiver operating characteristic (ROC) curves for the classification of age-related hearing loss based on GM modulated images or GM volume of the superior temporal gyrus; (a) and (b) correspond to training 1 and 2 of our CNN, (c) shows the classification performance of the logistic regression, and (d) the CNN performance with the sanity check. All classifications were based on a 5-fold cross validation. Abbreviations: CNN, convolutional neural network; AUC, Area under the curve.

D.2. Whole-brain classification

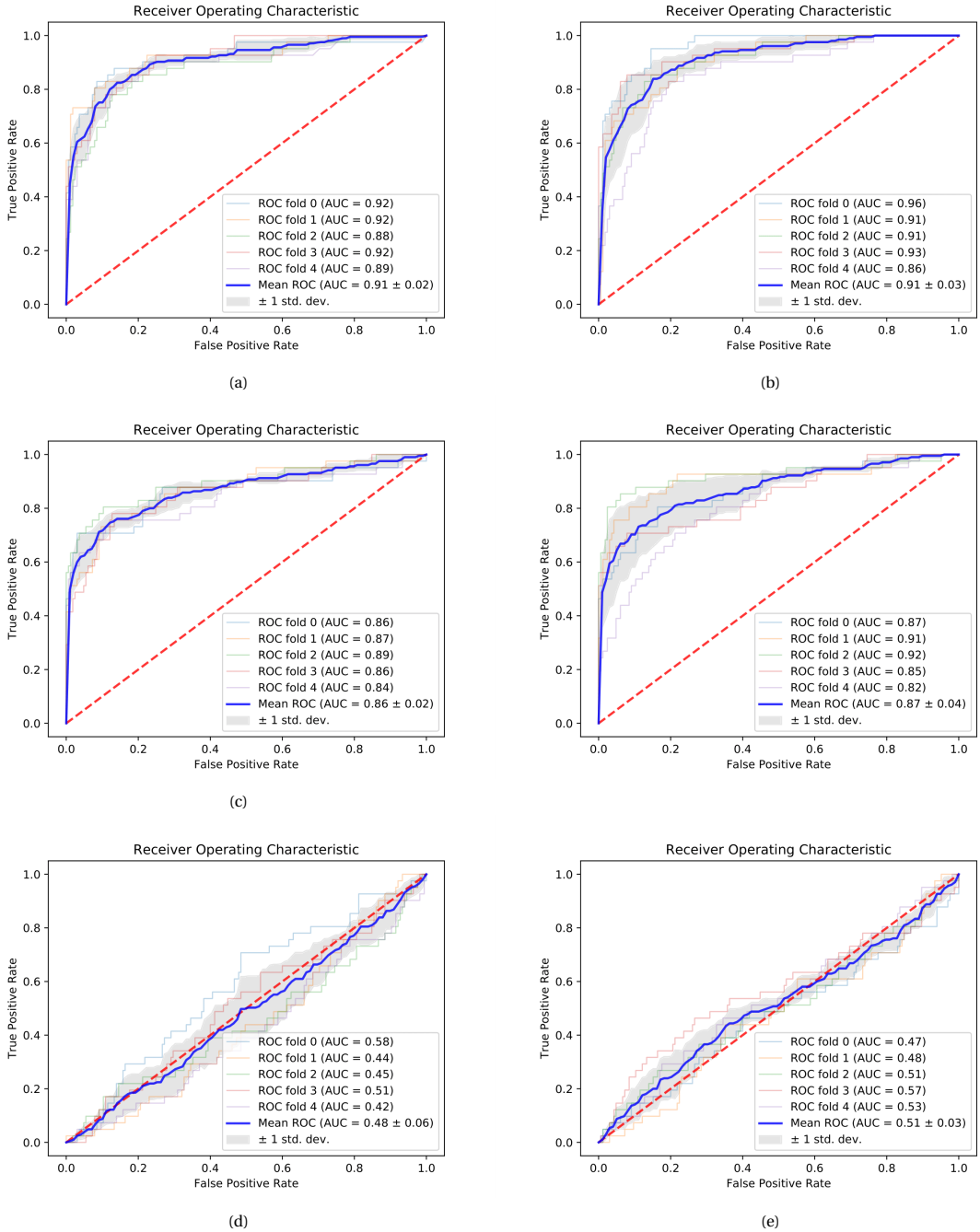
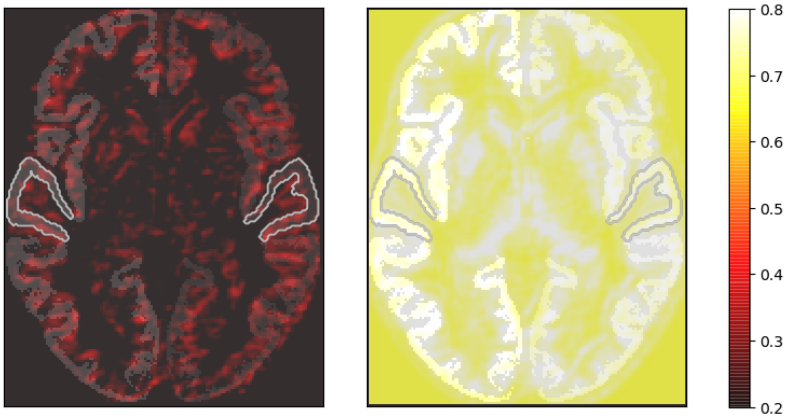


Figure D.2: Receiver operating characteristic (ROC) curves for the classification of age-related hearing loss based on GM modulated images of the whole brain; (a) and (b) are the CNN classification performances of training 1 and 2 based on MR-images without GM extraction, (c) and (d) with GM extraction, (d) and (e) their corresponding sanity checks. All classifications were based on a 5-fold cross validation. Abbreviations: CNN, convolutional neural network; AUC, Area under the curve.

E

Additional visualization results

E.1. Whole-brain images without GM extraction

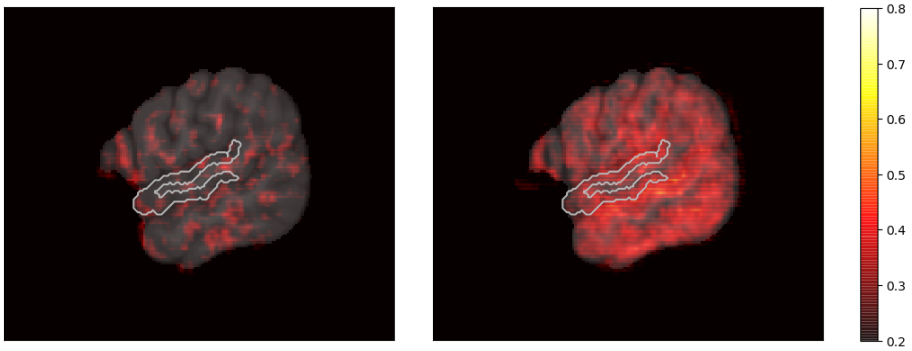


(a) (b)

Figure E.1: Visualization of the average network activation in the whole brain based on MR-images without GM extraction, axial view; (a) activation of the hearing loss class, (b) activation of the normal-hearing control class. The white matter contour corresponds to the superior temporal gyrus, obtained from the Desikan Killiany atlas [13]. The color bar shows the normalized intensity values of the Grad-CAM image.

E.2. Whole-brain activation

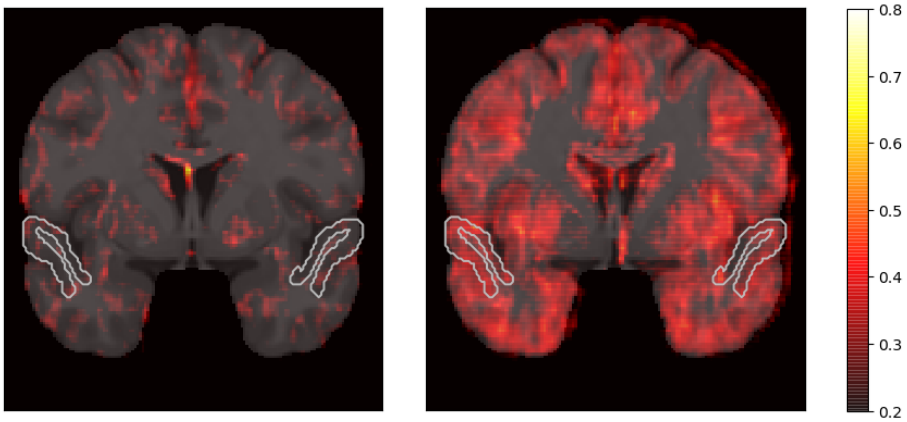
E.2.1. Sagittal view



(a) (b)

Figure E.2: Visualization of the average network activation in the whole brain based on MR-images with GM extraction, sagittal view; (a) activation of the hearing loss class, (b) activation of the normal-hearing control class. The white matter contour corresponds to the superior temporal gyrus, obtained from the Desikan Killiany atlas [13]. The colorbar shows the normalized intensity values of the Grad-CAM image.

E.2.2. Coronal view

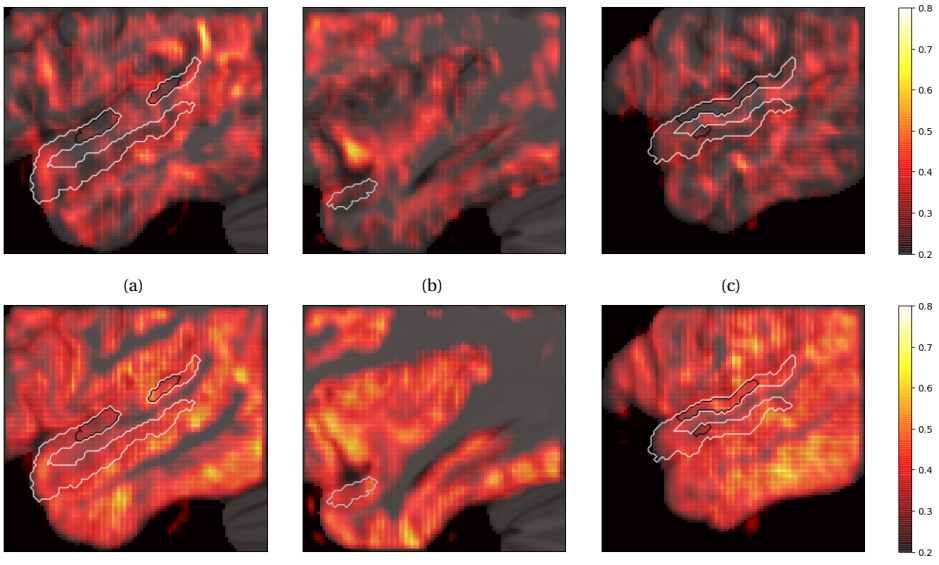


(a) (b)

Figure E.3: Visualization of the average network activation in the whole brain based on MR-images with GM extraction, coronal view; (a) activation of the hearing loss class, (b) activation of the normal-hearing control class. The white matter contour corresponds to the superior temporal gyrus, obtained from the Desikan Killiany atlas [13]. The colorbar shows the normalized intensity values of the Grad-CAM image.

E.3. Activation in the superior temporal gyrus region

E.3.1. Sagittal view



(a) (b) (c)

(d) (e) (f)

Figure E.4: Visualization of the average network activation in the superior temporal gyrus region, sagittal view; (a), (b) and (c) represent the hearing loss class at three positions; (d), (e) and (f) correspond to the activations of the hearing control class at the same three positions. The white matter contour corresponds to the superior temporal gyrus, obtained from the Desikan Killiany atlas [13]; the the black contour represents the primary auditory cortex, obtained from the Juelich atlas [19, 57]. The colorbar shows the normalized intensity values of the Grad-CAM image.

E.3.2. Coronal view

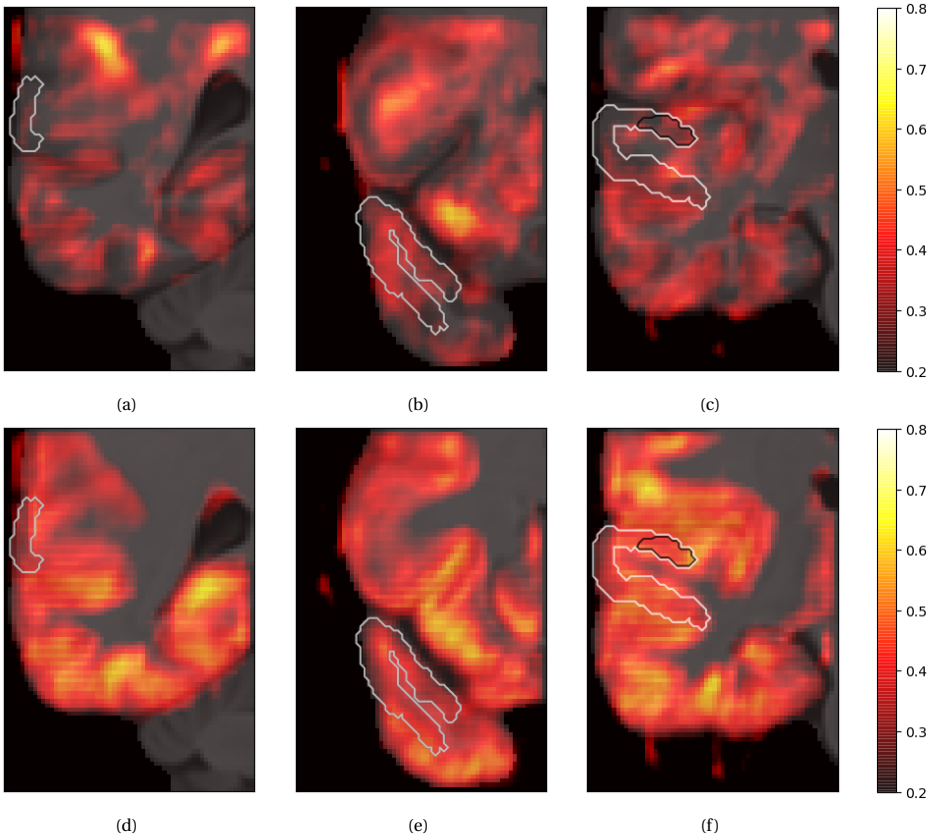
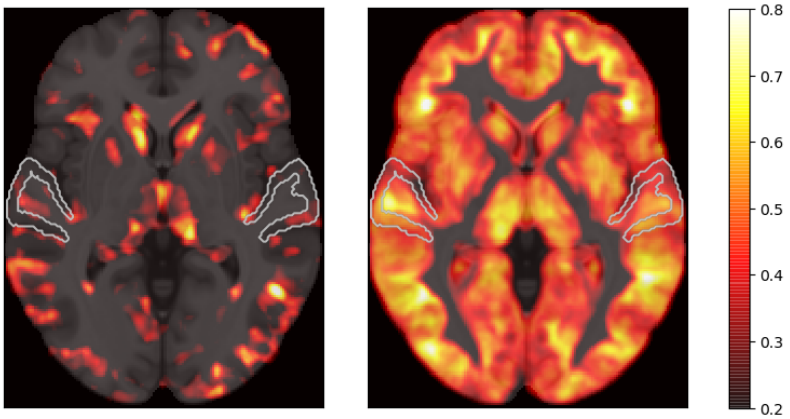


Figure E.5: Visualization of the average network activation in the superior temporal gyrus region, coronal view; (a), (b) and (c) represent the hearing loss class at three positions; (d), (e) and (f) correspond to the activations of the hearing control class at the same three positions. The white matter contour corresponds to the superior temporal gyrus, obtained from the Desikan Killiany atlas [13]; the the black contour represents the primary auditory cortex, obtained from the Juelich atlas [19, 57]. The colorbar shows the normalized intensity values of the Grad-CAM image.

E.4. Sanity check

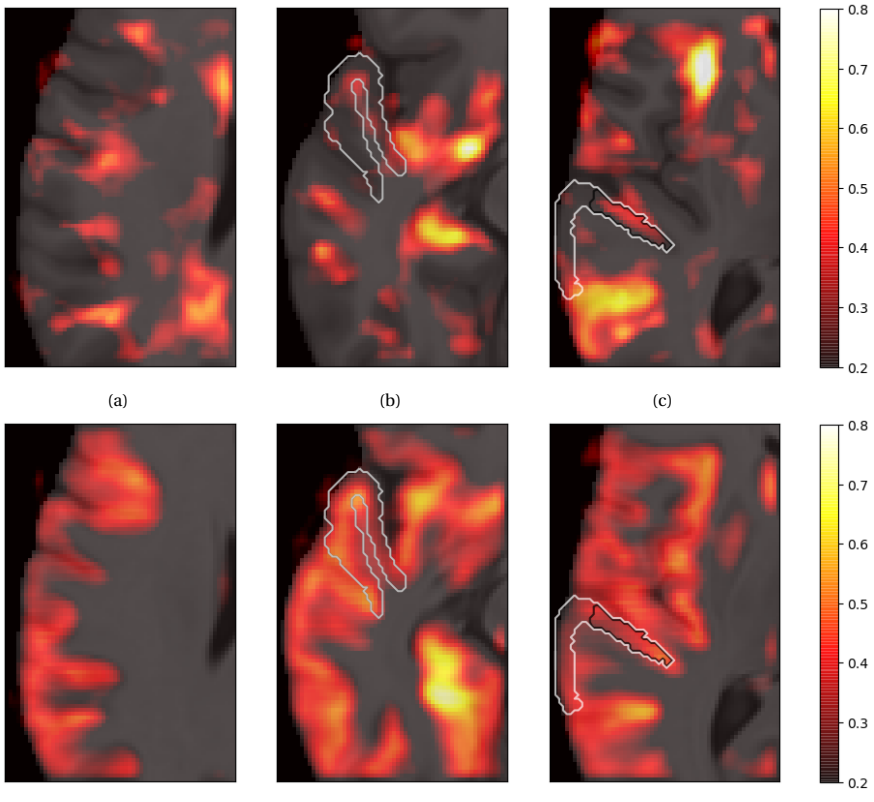
E.4.1. Whole brain



(a) (b)

Figure E.6: Visualization of the average network activation in the whole brain with the sanity check, axial view; (a) activation of the hearing loss class, (b) activation of the normal-hearing control class. The white matter contour corresponds to the superior temporal gyrus, obtained from the Desikan Killiany atlas [13]. The colorbar shows the normalized intensity values of the Grad-CAM image.

E.4.2. Superior temporal gyrus region



(a) (b) (c)

(d) (e) (f)

Figure E.7: Visualization of the average network activation in the superior temporal gyrus region with the sanity check, axial view; (a), (b) and (c) represent the hearing loss class at three positions; (d), (e) and (f) correspond to the activations of the hearing control class at the same three positions. The white matter contour corresponds to the superior temporal gyrus, obtained from the Desikan Killiany atlas [13]; the the black contour represents the primary auditory cortex, obtained from the Juelich atlas [19, 57]. The colorbar shows the normalized intensity values of the Grad-CAM image.

Bibliography

[1] Hakim C Achterberg, Lauge Sørensen, Frank J Wolters, Wiro J Niessen, Meike W Vernooij, M Arfan Ikram, Mads Nielsen, and Marleen de Bruijne. The value of hippocampal volume, shape, and texture for 11-year prediction of dementia: a population-based study. *Neurobiology of Aging*, 81:58–66, 2019. ISSN 0197-4580. doi: https://doi.org/10.1016/j.neurobiolaging.2019.05.007. URL http://www.sciencedirect.com/science/article/pii/S0197458019301423.

[2] Julius Adebayo, Justin Gilmer, Michael Muelly, Ian Goodfellow, Moritz Hardt, and Been Kim. Sanity checks for saliency maps. In *Advances in Neural Information Processing Systems*, pages 9505–9515, 2018.

[3] Jeffrey R Binder, Julie A Frost, Thomas A Hammeke, Robert W Cox, Stephen M Rao, and Thomas Prieto. Human Brain Language Areas Identified by Functional Magnetic Resonance Imaging. *The Journal of Neuroscience*, 17(1):353 LP – 362, 1 1997. doi: 10.1523/JNEUROSCI.17-01-00353.1997. URL http://www.jneurosci.org/content/17/1/353.abstract.

[4] Kris Boyen, Dave R M Langers, Emile de Kleine, and Pim van Dijk. Gray matter in the brain: Differences associated with tinnitus and hearing loss. *Hearing Research*, 295:67–78, 2013. ISSN 0378-5955. doi: https://doi.org/10.1016/j.heares.2012.02.010. URL http://www.sciencedirect.com/science/article/pii/S0378595512000469.

[5] Andreas Buja, Werner Stuetzle, and Yi Shen. Loss functions for binary class probability estimation and classification: Structure and applications. *Working draft, November*, 3, 2005.

[6] Velia Cardin. Effects of Aging and Adult-Onset Hearing Loss on Cortical Auditory Regions. *Frontiers in neuroscience*, 10:199, 5 2016. ISSN 1662-4548. doi: 10.3389/fnins.2016.00199. URL https://www.ncbi.nlm.nih.gov/pubmed/27242405https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4862970/.

[7] F Chollet. Deep Learning with Python, vol. 1. *Greenwich, CT: Manning Publications CO*, 2017.

[8] Kelly P Cosgrove, Carolyn M Mazure, and Julie K Staley. Evolving Knowledge of Sex Differences in Brain Structure, Function, and Chemistry. *Biological Psychiatry*, 62(8):847–855, 2007. ISSN 0006-3223. doi: https://doi.org/10.1016/j.biopsych.2007.03.001. URL http://www.sciencedirect.com/science/article/pii/S0006322307001989.

[9] Ruoxuan Cui and Manhua Liu. RNN-based longitudinal analysis for diagnosis of Alzheimer's disease. *Computerized Medical Imaging and Graphics*, 73:1–10, 2019. ISSN 0895-6111. doi: https://doi.org/10.1016/j.compmedimag.2019.01.005. URL http://www.sciencedirect.com/science/article/pii/S0895611118303987.

[10] BUREAU INTERNATIONAL D'AUDIOPHONOLOGIE. AUDIOMETRISCHE CLASSIFICATIE VAN GEHOORSTOORNISSEN, 1997. URL http://biap.org/index.php?option=com_content&view=article&id=67%3A-nederlandstalige-versie-aanbeveling-021-bis&catid=65%3Act-2-classification-des-surdites&Itemid=19&lang=en.

[11] Marleen de Bruijne. Machine learning approaches in medical image analysis: From detection to diagnosis. *Medical Image Analysis*, 33:94–97, 2016. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2016.06.032. URL http://www.sciencedirect.com/science/article/pii/S1361841516301098.

[12] Jennifer A Deal, Josh Betz, Kristine Yaffe, Tamara Harris, Elizabeth Purchase-Helzner, Suzanne Satterfield, Sheila Pratt, Nandini Govil, Eleanor M Simonsick, Frank R Lin, and for the Health A B C Study Group. Hearing Impairment and Incident Dementia and Cognitive Decline in Older Adults: The Health ABC Study. *The Journals of Gerontology: Series A*, 72(5):703–709, 4 2016. ISSN 1079-5006. doi: 10.1093/gerona/glw069. URL https://doi.org/10.1093/gerona/glw069.

[13] Rahul S. Desikan, Florent Ségonne, Bruce Fischl, Brian T. Quinn, Bradford C. Dickerson, Deborah Blacker, Randy L. Buckner, Anders M. Dale, R. Paul Maguire, Bradley T. Hyman, Marilyn S. Albert, and Ronald J. Killiany. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3):968–980, 7 2006. ISSN 1053-8119. doi: 10.1016/J.NEUROIMAGE.2006.01.021. URL https://www.sciencedirect.com/science/article/pii/S1053811906000437#bib19.

[14] Judy R Dubno, Mark A Eckert, Fu-Shing Lee, Lois J Matthews, and Richard A Schmiedt. Classifying human audiometric phenotypes of age-related hearing loss from animal models. *Journal of the Association for Research in Otolaryngology : JARO*, 14(5):687–701, 10 2013. ISSN 1438-7573. doi: 10.1007/s10162-013-0396-x. URL https://www.ncbi.nlm.nih.gov/pubmed/23740184https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3767874/.

[15] Florian Dubost, Gerda Bortsova, Hieab Adams, Arfan Ikram, Wiro Niessen, Meike Vernooij, and Marleen De Bruijne. GP-Unet: Lesion Detection from Weak Labels with a 3D Regression Network. 5 2017. URL http://arxiv.org/abs/1705.07999.

[16] Florian Dubost, Hieab Adams, Pinar Yilmaz, Gerda Bortsova, Gijs van Tulder, M Arfan Ikram, Wiro Niessen, Meike Vernooij, and Marleen de Bruijne. Weakly Supervised Object Detection with 2D and 3D Regression Neural Networks. *arXiv preprint arXiv:1906.01891*, 2019.

[17] Juergen Dukart, Matthias L Schroeter, Karsten Mueller, and The Alzheimer's Disease Neuroimaging Initiative. Age Correction in Dementia – Matching to a Healthy Brain. *PLOS ONE*, 6(7):e22193, 7 2011. URL https://doi.org/10.1371/journal.pone.0022193.

[18] Mark A Eckert, Stephanie L Cute, Kenneth I Vaden Jr, Stefanie E Kuchinsky, and Judy R Dubno. Auditory cortex signs of age-related hearing loss. *Journal of the Association for Research in Otolaryngology : JARO*, 13(5):703–713, 10 2012. ISSN 1438-7573. doi: 10.1007/s10162-012-0332-5. URL https://www.ncbi.nlm.nih.gov/pubmed/22618352https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3441956/.

[19] Simon B Eickhoff, Klaas E Stephan, Hartmut Mohlberg, Christian Grefkes, Gereon R Fink, Katrin Amunts, and Karl Zilles. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage*, 25(4):1325–1335, 2005. ISSN 1053-8119.

[20] Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 6 2006. ISSN 0167-8655. doi: 10.1016/J.PATREC.2005.10.010. URL https://www.sciencedirect.com/science/article/pii/S016786550500303X.

[21] Bruce Fischl, Martin I Sereno, and Anders M Dale. Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. *Neuroimage*, 9(2):195–207, 1999. ISSN 1053-8119.

[22] V S Fonov, A C Evans, R C McKinstry, C R Almli, and D L Collins. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, 47:S102, 2009. ISSN 1053-8119. doi: https://doi.org/10.1016/S1053-8119(09)70884-5. URL http://www.sciencedirect.com/science/article/pii/S1053811909708845.

[23] S Fortunato, F Forli, V Guglielmi, E De Corso, G Paludetti, S Berrettini, and A R Fetoni. A review of new insights on the association between hearing loss and cognitive decline in ageing. *Acta otorhinolaryngologica Italica : organo ufficiale della Societa italiana di otorinolaringologia e chirurgia cervico-facciale*, 36(3):155–166, 6 2016. ISSN 1827-675X. doi: 10.14639/0392-100X-993. URL https://www.ncbi.nlm.nih.gov/pubmed/27214827https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4977003/.

[24] Erik Fransen, Vedat Topsakal, Jan-Jaap Hendrickx, Lut Van Laer, Jeroen R Huyghe, Els Van Eyken, Nele Lemkens, Samuli Hannula, Elina Mäki-Torkko, Mona Jensen, Kelly Demeester, Anke Tropitzsch, Amanda Bonaconsa, Manuela Mazzoli, Angeles Espeso, Katia Verbruggen, Joke Huyghe, Patrick L M Huygen, Sylvia Kunst, Minna Manninen, Amalia Diaz-Lacava, Michael Steffens, Thomas F Wienker, Ilmari Pyykkö, Cor W R J Cremers, Hannie Kremer, Ingeborg Dhooge, Dafydd Stephens, Eva Orzan, Markus Pfister, Michael Bille, Agnete Parving, Martti Sorri, Paul Van de Heyning, and Guy Van Camp. Occupational noise, smoking, and a high body mass index are risk factors for age-related hearing impairment and

moderate alcohol consumption is protective: a European population-based multicenter study. *Journal of the Association for Research in Otolaryngology : JARO*, 9(3):263–264, 9 2008. ISSN 1525-3961. doi: 10.1007/s10162-008-0123-1. URL https://www.ncbi.nlm.nih.gov/pubmed/18543032https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2492985/.

[25] George A. Gates and John H. Mills. Presbycusis. In *Lancet*, volume 366, pages 1111–1120, 9 2005. doi: 10.1016/S0140-6736(05)67423-5.

[26] Aurélien Géron. *Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems*. " O'Reilly Media, Inc.", 2017. ISBN 1491962267.

[27] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256, 2010.

[28] Justin S Golub, José A Luchsinger, Jennifer J Manly, Yaakov Stern, Richard Mayeux, and Nicole Schupf. Observed Hearing Loss and Incident Dementia in a Multiethnic Cohort. *Journal of the American Geriatrics Society*, 65(8):1691–1697, 8 2017. ISSN 0002-8614. doi: 10.1111/jgs.14848. URL https://doi.org/10.1111/jgs.14848.

[29] Catriona D. Good, Ingrid S. Johnsrude, John Ashburner, Richard N.A. Henson, Karl J. Friston, and Richard S.J. Frackowiak. A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage*, 14(1 I):21–36, 2001. ISSN 10538119. doi: 10.1006/nimg.2001.0786.

[30] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016. ISBN 0262337371.

[31] Richard Klaus Gurgel, Preston Daniel Ward, Sarah Schwartz, Maria C Norton, Norman L Foster, and JoAnn T Tschanz. Relationship of hearing loss and dementia: a prospective, population-based study. *Otology & neurotology : official publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, 35(5):775–781, 6 2014. ISSN 1537-4505. doi: 10.1097/MAO.0000000000000313. URL https://www.ncbi.nlm.nih.gov/pubmed/24662628https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4024067/.

[32] Troy A. Hackett. *Anatomic organization of the auditory cortex*, volume 129. Elsevier B.V., 1 edition, 2015. doi: 10.1016/B978-0-444-62630-1.00002-0. URL http://dx.doi.org/10.1016/B978-0-444-62630-1.00002-0.

[33] Timothy C Hain. Chapter 12 - Cranial Nerve VIII: Vestibulocochlear System. pages 199–215. W.B. Saunders, Philadelphia, 2007. ISBN 978-1-4160-3618-0. doi: https://doi.org/10.1016/B978-141603618-0.10012-8. URL http://www.sciencedirect.com/science/article/pii/B9781416036180100128.

[34] Thomas J Hoffmann, Bronya J Keats, Noriko Yoshikawa, Catherine Schaefer, Neil Risch, and Lawrence R Lustig. A Large Genome-Wide Association Study of Age-Related Hearing Impairment Using Electronic Health Records. *PLOS Genetics*, 12(10):e1006371, 10 2016. URL https://doi.org/10.1371/journal.pgen.1006371.

[35] A Hofman, D E Grobbee, P T V M De Jong, and F A Van den Ouweland. Determinants of disease and disability in the elderly: The Rotterdam elderly study. *European Journal of Epidemiology*, 7(4):403–422, 1991. ISSN 1573-7284. doi: 10.1007/BF00145007. URL https://doi.org/10.1007/BF00145007.

[36] J D Hood. The principles and practice of bone conduction audiometry: A review of the present position. *The Laryngoscope*, 70(9):1211–1228, 1960. ISSN 0023-852X.

[37] A Howarth and G R Shone. Ageing and the auditory system. *Postgraduate Medical Journal*, 82(965):166 LP – 171, 3 2006. doi: 10.1136/pgmj.2005.039388. URL http://pmj.bmj.com/content/82/965/166.abstract.

[38] W Huizinga, D H J Poot, M W Vernooij, G V Roshchupkin, E E Bron, M A Ikram, D Rueckert, W J Niessen, and S Klein. A spatio-temporal reference model of the aging brain. *NeuroImage*, 169:11–22, 2018. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2017.10.040. URL http://www.sciencedirect.com/science/article/pii/S1053811917308674.

[39] Larry E Humes, Thomas A Busey, James Craig, and Diane Kewley-Port. Are age-related changes in cognitive function driven by age-related changes in sensory processing? *Attention, Perception, & Psychophysics*, 75(3):508–524, 2013. ISSN 1943-393X. doi: 10.3758/s13414-012-0406-9. URL https://doi.org/10.3758/s13414-012-0406-9.

[40] M Arfan Ikram, Aad van der Lugt, Wiro J Niessen, Peter J Koudstaal, Gabriel P Krestin, Albert Hofman, Daniel Bos, and Meike W Vernooij. The Rotterdam Scan Study: design update 2016 and main findings. *European journal of epidemiology*, 30(12):1299–1315, 12 2015. ISSN 1573-7284. doi: 10.1007/s10654-015-0105-7. URL https://www.ncbi.nlm.nih.gov/pubmed/26650042https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4690838/.

[41] M Arfan Ikram, Guy G O Brusselle, Sarwa Darwish Murad, Cornelia M van Duijn, Oscar H Franco, André Goedegebure, Caroline C W Klaver, Tamar E C Nijsten, Robin P Peeters, Bruno H Stricker, Henning Tiemeier, André G Uitterlinden, Meike W Vernooij, and Albert Hofman. The Rotterdam Study: 2018 update on objectives, design and main results. *European journal of epidemiology*, 32(9):807–850, 9 2017. ISSN 1573-7284. doi: 10.1007/s10654-017-0321-4. URL https://www.ncbi.nlm.nih.gov/pubmed/29064009https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5662692/.

[42] International Organization for Standardization. ISO 8253-1:2010. Acoustics—audiometric test methods—Part 1: basic pure-tone air and bone conduction threshold audiometry. ISO 8253-1:2010).

[43] Sergey Ioffe and Christian Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. 2 2015. URL http://arxiv.org/abs/1502.03167.

[44] Dona M P Jayakody, Osvaldo P Almeida, Craig P Speelman, Rebecca J Bennett, Thomas C Moyle, Jessica M Yiannos, and Peter L Friedland. Association between speech and high-frequency hearing loss and depression, anxiety and stress in older adults. *Maturitas*, 110:86–91, 4 2018. ISSN 0378-5122. doi: 10.1016/j.maturitas.2018.02.002. URL https://doi.org/10.1016/j.maturitas.2018.02.002.

[45] Dona M P Jayakody, Peter L Friedland, Ralph N Martins, and Hamid R Sohrabi. Impact of Aging on the Auditory System and Related Cognitive Functions: A Narrative Review. *Frontiers in neuroscience*, 12:125, 3 2018. ISSN 1662-4548. doi: 10.3389/fnins.2018.00125. URL https://www.ncbi.nlm.nih.gov/pubmed/29556173https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5844959/.

[46] J.Linders. CNN for AD classification, 2019. URL https://github.com/Jara555/CNN-for-AD-classification.

[47] J. Johnson and A. Karpathy. CS231n Convolutional Neural Networks for Visual Recognition, 2019. URL http://cs231n.github.io/.

[48] Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pages 1–15, 2015.

[49] Robert A Levine and Yahav Oron. Tinnitus. In *Handbook of clinical neurology*, volume 129, pages 409–431. Elsevier, 2015. ISBN 0072-9752.

[50] F R Lin, L Ferrucci, Y An, J O Goh, Jimit Doshi, E J Metter, C Davatzikos, M A Kraut, and S M Resnick. Association of hearing impairment with brain volume changes in older adults. *NeuroImage*, 90:84–92, 2014. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2013.12.059. URL http://www.sciencedirect.com/science/article/pii/S1053811914000032.

[51] Frank R Lin, E Jeffrey Metter, Richard J O'Brien, Susan M Resnick, Alan B Zonderman, and Luigi Ferrucci. Hearing Loss and Incident Dementia. *JAMA Neurology*, 68(2):214–220, 2 2011. ISSN 2168-6149. doi: 10.1001/archneurol.2010.362. URL https://doi.org/10.1001/archneurol.2010.362.

[52] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[53] Ulman Lindenberger and Paul B Baltes. Sensory functioning and intelligence in old age: A strong connection., 1994.

[54] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A W M van der Laak, Bram van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2017.07.005. URL http://www.sciencedirect.com/science/article/pii/S1361841517301135.

[55] P. McCarthy. FNIRT/UserGuide, 2018. URL https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FNIRT/UserGuide.

[56] Paul Mick, Ichiro Kawachi, and Frank R Lin. The Association between Hearing Loss and Social Isolation in Older Adults. *Otolaryngology–Head and Neck Surgery*, 150(3):378–384, 1 2014. ISSN 0194-5998. doi: 10.1177/0194599813518021. URL https://doi.org/10.1177/0194599813518021.

[57] P Morosan, J Rademacher, A Schleicher, K Amunts, T Schormann, and K Zilles. Human Primary Auditory Cortex: Cytoarchitectonic Subdivisions and Mapping into a Spatial Reference System. *NeuroImage*, 13 (4):684–701, 2001. ISSN 1053-8119. doi: https://doi.org/10.1006/nimg.2000.0715. URL http://www.sciencedirect.com/science/article/pii/S1053811900907158.

[58] In-Hwan Oh, Jong Hoon Lee, Dong Choon Park, MyungGu Kim, Ji Hyun Chung, Sang Hoon Kim, and Seung Geun Yeo. Hearing Loss as a Function of Aging and Diabetes Mellitus: A Cross Sectional Study. *PLOS ONE*, 9(12):e116161, 12 2015. URL https://doi.org/10.1371/journal.pone.0116161.

[59] Jonathan E Peelle, Vanessa Troiani, Murray Grossman, and Arthur Wingfield. Hearing loss in older adults affects neural systems supporting speech comprehension. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(35):12638–12643, 8 2011. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.2559-11.2011. URL https://www.ncbi.nlm.nih.gov/pubmed/21880924https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3175595/.

[60] M Kathleen Pichora-Fuller and Pamela E Souza. Effects of aging on auditory processing of speech. *International journal of audiology*, 42(sup2):11–16, 2003. ISSN 1499-2027.

[61] Saima Rathore, Mohamad Habes, Muhammad Aksam Iftikhar, Amanda Shacklett, and Christos Davatzikos. A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *NeuroImage*, 155:530–548, 7 2017. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2017.03.057. URL https://www.ncbi.nlm.nih.gov/pubmed/28414186https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5511557/.

[62] Fuxin Ren, Wen Ma, Muwei Li, Huaiqiang Sun, Qian Xin, Wei Zong, Weibo Chen, Guangbin Wang, Fei Gao, and Bin Zhao. Gray Matter Atrophy Is Associated With Cognitive Impairment in Patients With Presbycusis: A Comprehensive Morphometric Study. *Frontiers in neuroscience*, 12:744, 10 2018. ISSN 1662-4548. doi: 10.3389/fnins.2018.00744. URL https://www.ncbi.nlm.nih.gov/pubmed/30405333https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6205975/.

[63] Susan M Resnick, Dzung L Pham, Michael A Kraut, Alan B Zonderman, and Christos Davatzikos. Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 23(8):3295–3301, 4 2003. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.23-08-03295.2003. URL https://www.ncbi.nlm.nih.gov/pubmed/12716936https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6742337/.

[64] Stephanie C. Rigters, Mick Metselaar, Marjan H. Wieringa, Robert J.Baatenburg De Jong, Albert Hofman, and André Goedegebure. Contributing Determinants to Hearing Loss in Elderly Men and Women: Results from the Population-Based Rotterdam Study. *Audiology and Neurotology*, 21(1):10–15, 2016. ISSN 14219700. doi: 10.1159/000448348.

[65] Stephanie C. Rigters, Daniel Bos, Mick Metselaar, Gennady V. Roshchupkin, Robert J. Baatenburg de Jong, M. Arfan Ikram, Meike W. Vernooij, and André Goedegebure. Hearing impairment is associated with smaller brain volume in aging. *Frontiers in Aging Neuroscience*, 9(JAN), 1 2017. ISSN 16634365. doi: 10.3389/fnagi.2017.00002.

[66] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019. ISSN 2522-5839.

[67] Graeme D Ruxton. The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behavioral Ecology*, 17(4):688–690, 5 2006. ISSN 1045-2249. doi: 10.1093/beheco/ark016. URL https://doi.org/10.1093/beheco/ark016.

[68] R A Schmiedt. The aging auditory system. 2010.

[69] Bruce A Schneider and M Kathleen Pichora-Fuller. Implications of perceptual deterioration for cognitive aging research. In *The handbook of aging and cognition, 2nd ed.*, pages 155–219. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 2000. ISBN 0-8058-2966-0 (Hardcover).

[70] Harold F Schuknecht and Mark R Gacek. Cochlear Pathology in Presbycusis. *Annals of Otology, Rhinology & Laryngology*, 102(1_suppl):1–16, 1 1993. ISSN 0003-4894. doi: 10.1177/00034894931020S101. URL https://doi.org/10.1177/00034894931020S101.

[71] Florent Ségonne, Anders M Dale, Evelina Busa, Maureen Glessner, David Salat, Horst K Hahn, and Bruce Fischl. A hybrid approach to the skull stripping problem in MRI. *Neuroimage*, 22(3):1060–1075, 2004. ISSN 1053-8119.

[72] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 618–626, 2017.

[73] John G Sled, Alex P Zijdenbos, and Alan C Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE transactions on medical imaging*, 17(1):87–97, 1998. ISSN 0278-0062.

[74] Henri A. Vrooman, Chris A. Cocosco, Fedde van der Lijn, Rik Stokking, M. Arfan Ikram, Meike W. Vernooij, Monique M.B. Breteler, and Wiro J. Niessen. Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification. *NeuroImage*, 37(1):71–81, 8 2007. ISSN 1053-8119. doi: 10.1016/J.NEUROIMAGE.2007.05.018. URL https://www.sciencedirect.com/science/article/pii/S1053811907004181.

[75] Rachel V Wayne and Ingrid S Johnsrude. A review of causal mechanisms underlying the link between age-related hearing loss and cognitive decline. *Ageing Research Reviews*, 23:154–166, 2015. ISSN 1568-1637. doi: https://doi.org/10.1016/j.arr.2015.06.002. URL http://www.sciencedirect.com/science/article/pii/S1568163715000707.

[76] Bernard L Welch. The generalization ofstudent's' problem when several different population variances are involved. *Biometrika*, 34(1/2):28–35, 1947. ISSN 0006-3444.

[77] World Health Organisation. World Health Organisation Global Estimates on Prevalence of Hearing Loss. pages 1–15, 2012. doi: 10.1002/2014GB005021.

[78] Wei Xu, Can Zhang, Jie-Qiong Li, Chen-Chen Tan, Xi-Peng Cao, Lan Tan, Jin-Tai Yu, and Alzheimer's Disease Neuroimaging Initiative. Age-related hearing loss accelerates cerebrospinal fluid tau levels and brain atrophy: a longitudinal study. *Aging*, 11(10):3156–3169, 5 2019. ISSN 1945-4589. doi: 10.18632/aging.101971. URL https://www.ncbi.nlm.nih.gov/pubmed/31118310https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6555452/.