# Exploratory data mining to examine the impact of internet-based cognitive behavioral therapy for tinnitus: Application of decision tree models

**Hansapani Rodrigo,[1] Eldré W. Beukes,[2,3] Gerhard Andersson,[4,5] & Vinaya Manchaiah,[2,6]**

1. School of Mathematical and Statistical Sciences, University of Texas Rio Grande Valley, Texas, USA
2. Department of Speech and Hearing Sciences, Lamar University, Beaumont, Texas, USA
3. Department of Vision and Hearing Sciences, School of Psychology and Sport Science, Anglia Ruskin University, Cambridge, United Kingdom
4. Department of Behavioral Sciences and Learning, Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden
5. Department of Clinical Neuroscience, Division of Psychiatry, Karolinska Institute, Stockholm, Sweden
6. Department of Speech and Hearing, School of Allied Health Sciences, Manipal, Karnataka, India

## Corresponding author:

Hansapani Rodrigo
University of Texas Rio Grande Valley
School of Mathematical and Statistical Sciences
1201, W University Drive,
Edinburg, TX, 78539
United States
Phone: 1 956-665-2349
Email: hansapani.rodrigo@utrgv.edu

## Abstract

**Background:**
There is a huge variability in the way individuals with tinnitus respond to interventions. These experiential variations together with a range of associated etiologies, contribute to tinnitus being a highly heterogeneous condition. Despite this heterogeneity, a “one size fits all” approach is taken when making management recommendations. Although there are various management approaches, not all are equally effective. Psychological approaches such as cognitive behavioral therapy (CBT) have the most evidence-base.

**Objectives:** Managing tinnitus is challenging due to the significant variations in tinnitus experiences and treatment success. Tailored interventions based on individual tinnitus profiles may improve outcomes. Predictive models of treatment success are, however, lacking. The current study aimed to used exploratory data mining techniques (i.e., decision tree models) to identify the variables associated with treatment success for an Internet-based cognitive behavioral therapy (ICBT) for tinnitus.

**Methods:** Individuals (n = 228) who underwent ICBT in three separate clinical trials were included in this analysis. The primary outcome variable was reducing 13 points in tinnitus severity as measured by the Tinnitus Functional Index following the intervention. Predictor variables included demographic characteristics, tinnitus, and hearing-related variables, and clinical factors (i.e., anxiety, depression, insomnia, hyperacusis, hearing disability, cognitive function, and life satisfaction). Analyses were undertaken using various exploratory machine learning algorithms to identify the most suitable variable. Six decision tree models were implemented, namely Classification and decision trees (CART), C5.0, Gradient Boosting, AdaBoost algorithm, eXtreme Gradient Boosting and Random Forest. The SHapley Additive exPlanations (SHAP) framework was applied to the two best models to identify the relative predictor importance.

**Results:** Of the six decision tree models, CART [accuracy of 70.7% (SD 2.4) sensitivity of 74.0% (SD 5.5), specificity of 64% (SD 3.7), and area under the receiver operating characteristic curve (AUC) $0.69 \pm 0.001$)] and Gradient boosting [accuracy of 71.8% (SD 1.5), sensitivity of 78.3% (SD 2.8), specificity of 58.7% (SD 4.2), and AUC 0.68 (SD 0.02) were found to be the best predictive models. Although the other models had an acceptable accuracy (ranged between 56.3 to 66.7%) and sensitivity (varied between 68.6 to 77.9%), they all had relatively weak specificity (varied between 31.1 to 50.0%) and AUC (varied between .52 to .62). Higher education level was the most influencing factors in the ICBT outcome. The CART decision tree model identified three participant groups who had at least 85% success probability following undertaking ICBT.

**Conclusions:** In this study, decision tree models, especially the CART and Gradient Boosting models, appear to be promising in predicting the ICBT outcomes. Their predictive power may be improved by using larger sample sizes and including a wider range of predictive factors in future studies.

**Registration:** Clinicaltrials.gov NCT02370810; Clinicaltrials.gov NCT02665975.

## Key Words

Tinnitus, Internet interventions, Digital therapeutics, Cognitive behavioral therapy, Artificial intelligence, Machine learning, Data mining, Decision tree, Random forest

## Introduction

### Background

Tinnitus is the perception of a sound in the ears or head in the absence of a corresponding external sound source. It is very prevalent, with an estimated 10-15% of the adult population experiencing tinnitus [1]  Various conditions are associated with developing tinnitus, such as ear disorders [2] exposure to loud noise, presence of a hearing loss, and increasing age [3] Tinnitus experiences are highly heterogeneous in terms of how it is manifested (e.g., types of sounds experienced, how individuals react to their sounds, the associated comorbidities,) and also how individuals with tinnitus respond to treatment [4] Although a majority of those with tinnitus are not bothered by their tinnitus, a significant number experience distressing tinnitus that affects their quality of life [5] Although tinnitus can affect people in different ways, the most complaints include annoyance, irritability, fatigue, stress, sleep problems, trouble concentrating [6] Moreover, distressing tinnitus is often associated with an increased risk of anxiety and depression [7,8].Various management strategies are used to help persons with tinnitus, including sound therapy (e.g., hearing aids, masking), informational counseling to aid understanding of tinnitus, psychological approaches addressing unhelpful thought patterns and reactions to tinnitus such as Cognitive Behavioural Therapy (CBT). Of these, CBT has the highest level of research evidence in reducing tinnitus distress [9,10].

Although the use of CBT is recommended in many tinnitus practice guidelines [11], it is seldomly provided, partly due to a lack of trained professionals who can offer CBT for tinnitus in an in-person format. To overcome this barrier, internet-based CBT (ICBT) was developed in the late 1990's [12]. In ICBT, the treatment strategies are offered to individuals with tinnitus as self-help materials over the Internet together with professional guidance [13]. The feasibility and efficacy of such an approach have been demonstrated in several populations in Sweden, Germany, Australia, and the UK [14], and more recently in the US [15,16]. In general, the studies have shown that nearly 50-60% of those who undergo ICBT will have a clinically significant reduction of tinnitus distress [17,18]. To date, no strong predictors of ICBT outcome have been identified to indicate who is likely to benefit from ICBT interventions. Predictors of outcome identified when examining the long-term (1-year) outcomes of ICBT in the UK were higher baseline tinnitus severity, more engagement with the ICBT program (i.e., more modules opened), and higher self-reported satisfaction with the intervention [18]. To further explore predictors of outcome, various univariate and multivariate (i.e., logistic and linear) regression models were applied to a combined dataset of multiple ICBT studies [19]. The linear and logistic regression models have identified education level (linear: $P$=.01, logistic: $P$<.001), and baseline tinnitus severity (linear: $P$<.001, logistic: $P$<.001), to be significant predictor variables contributing to reduction in tinnitus severity post-ICBT intervention. As per linear regression model, participants who had received disability allowance had shown 25.30-point (95% C.I: -46.35,-4.24) lower Tinnitus Functional Index (TFI) reduction compared to those who didn't have to work less due to tinnitus after adjusting for baseline tinnitus severity and their education level. Although, many other predictors including age, tinnitus duration, loud noise exposure, etc [19] were not identified to be significant under these linear models, there is a possibility that those non-significant variables might be associated with the response, in a non-liner setting.

In the last two decades, various Artificial intelligence (AI) and/or Machine Learning (ML) techniques have been developed and applied to hearing health data. Such approaches have mainly been used in disease profiling, although some studies have focused on the prediction of treatment outcomes [20-24]. It is noteworthy that the intervention trials in audiology and also in tinnitus research usually involve a few hundred participants, and generally extensive data regarding demographic characteristics and clinical variables are collected. Such a data set with many predictor variables may be best handled by exploratory data mining techniques such as tree-based models like the random forest. Such models tend to perform well even in the presence of multicollinearity among a large number of predictor variables as they decorrelate the variables [25]. For example, a recent study that examined various ML algorithms in predicting the CBT treatment outcome in the tinnitus population suggested that gradient boosted trees (Area Under the receiver operating characteristic curve [AUC] of 0.89) had the best predictive power [21]. This study identified that subjectively perceived tinnitus-related impairment, depression, sleep problems, physical health-related impairments in quality of life, time spent to complete questionnaires, and educational level exhibited a high contribution towards model prediction.

However, no previous studies have examined the application of AI and/or ML techniques on ICBT outcomes in tinnitus research.

### Objectives

To further explore outcome predictors for ICBT, the objective of this study was to examine applications of various exploratory data mining techniques based on decision tree models. In particular, we wanted to investigate (a) which types of decision tree models were most applicable to ICBT outcome prediction (i.e., models with the best accuracy and predictive power), and (b) identify the most relevant predictive factors of ICBT outcomes using the most appropriate decision tree models.

## Methods

### Study Design and Ethical Considerations

We included 228 participants who previously underwent ICBT for tinnitus and whose data were collected as a part of three separate ICBT trials [17,18,26] during 2016-18. This study was the secondary analysis of these ICBT intervention studies.  Ethical clearance was obtained from the Faculty of Science and Technology Research Ethics Panel of Anglia Ruskin University (ARU reference: FST/FREP/14/478 and FST/FREP/14/478) and the East of England–Cambridge South Research Ethics Committee (REC reference: 16/EE/0148) and Health Research Authority (IRAS project ID: 195565).

### Participant Characteristics

A heterogenous tinnitus sample was obtained, thus representing typical tinnitus populations as seen in Appendix 1. The average age was 55.14 years (SD 12.92), and 98 out of 228 (43%) were females. The majority had long-standing tinnitus with a mean duration of 17.68 (SD 19.42) years. Out of 228, 59 (26%) had completed high school education, 61 (26%) had an undergraduate degree, and only 30 (13%) had a postgraduate degree. Approximately 48% (109 out of 228) of the participants had their tinnitus in both ears, while 61 (27%) of them had tinnitus had only in one ear, with others reporting tinnitus in their head or other locations. The majority, 159 out of 228 (70%) did not wear hearing aid/s, and 58 (25%) of them had some sought of tinnitus treatment previously.

### Intervention

The study participants completed an 8-week ICBT intervention which was presented in a self-help format [13,27]. The intervention was administered using a secured ePlatform [28,29]. During this 8-week period, participants were represented with 2-3 learning modules that contained various elements of CBT specifically adapted for tinnitus, including applied relaxation, cognitive restructuring, and imagery. The digital materials were presented using text, images, and videos. In addition, various exercises were presented in these learning modules to improve engagement.

### Data Collection

The baseline data collection included an extensive questionnaire that focused on demographics, tinnitus-related and treatment-related information. Outcome data were gathered using standardized primary and secondary self-reported questionnaires, which were administered before (baseline), during (weekly), and post-intervention. The primary outcome was a change in tinnitus severity, as measured by the Tinnitus Functional Index [30]. The *secondary outcome measures* included the Insomnia Severity Index (ISI; [31]) as a measure of insomnia, the Generalized Anxiety Disorder (GAD-7; [32]) as a measure of anxiety, the Patient Health Questionnaire (PHQ-9; [33]) as a measure of depressive symptoms, the Hearing Handicap Inventory for Adults Screening version (HHIA-S; [34]) as a measure of self-reported hearing disability, the Hyperacusis Questionnaire (HQ; [35]) to assess the presence hyperacusis (i.e., reduced tolerance of everyday sounds), the Cognitive Failures Questionnaire (CFQ; [36]) to assess cognitive functions, and the Satisfaction with Life Scales (SWLS; [37]) to assess global life satisfaction.

## Data Analyses

### *Variables*

The primary outcome (dependent) variable in this study was the change in tinnitus severity. A 13-point reduction in TFI scores following the ICBT intervention was regarded as a clinically significant (successful) treatment outcome [30]. Significant differences in scores were assessed using paired sample *t*-tests. Significance was set to $P = 0.05$. There were 33 predictor variables selected as outlined in Appendix 2. These included:

- Seven demographic variables (i.e., age, gender, education level, employment type, noise exposure, the presence of psychological conditions, tinnitus affecting the ability to work),
- Fifteen tinnitus and hearing-related variables (i.e., baseline tinnitus severity, tinnitus duration, how often tinnitus heard, tinnitus location,
- Nine different types of tinnitus types, multiple tones heard, and the presence of hearing loss),
- Four treatment-related variables (i.e., past treatment sought, tinnitus maskability, hearing aid use, and medication use), and
- Seven clinical factors (i.e., anxiety, depression, insomnia, hyperacusis, hearing disability, cognitive functions, and life satisfaction.

### *Decision Tree Models/Classifiers*

Data analysis focused on decision tree-based models as they play an essential role in exploratory data mining and facilitate human decision-making by providing decision rules [38]. Despite their simplicity, decision trees usually exhibit higher variance in their predictions and are not consistently robust. Given this, their powerful counterparts like Random Forest (RF;[39]), gradient boosting models (GB; [40]) and eXtreme Gradient Boosting (XGBoost;[41]) were selected. For comparison, six decision tree models were used, namely: Classification and Regression Tree (CART;[42]), C5.0 [43], GB, AdaBoost algorithm [44], and RF. As CART, C5.0, and RF decision tree models involve stratifying or segmenting the predictor space into a number of non-overlapping regions [38], recursive binary splitting for classification using Gini-index was used [45]. Many of these decision tree types have been applied on audiological data and found to have good results in previous studies [20-24].

### *Data Analyses Steps*

The analyses were performed in four stages. First, the data were split into training and testing data, and the classifier models were trained on a dataset before testing. Second, the six classifiers were applied to the data to identify the most suitable models based on their performance evaluation. Third, the two best models were further used to determine the predictors of ICBT outcomes. The steps are described in more detail below.

***Step 1 – Classifier Training:*** Prior to applying the decision tree classifiers, the entire data set was divided into training (80%, n = 183) and testing (20%, n = 45) data sets. The training data set was used to develop the corresponding data mining model while the testing data set was used to evaluate the model predictions. As the training data set was relatively small (n=183), a repeated 3-fold cross-validation was incorporated (except for the CART model where the full training data set has been used for model training). With this approach, each fold was given a chance to act as their own validation set to minimize the propensity of model overfitting. Ten different models were created with several random initializations for each data mining method. Hyper-parameter tunning for each of these decision models has been performed, as required. For instance, when training the RF models, we have explored a range of different number of predictors for splitting at each tree node and its impact on the model performances.

***Step 2 – Classifier Performance Evaluation:*** The trained models were evaluated using the testing data set in terms of their mean predictive accuracy, sensitivity (true positive rate), specificity (true negative rate), and the AUC. These are given as mean$\pm SD$ based on the 10 replicated models for each data mining technique. AUC is used as a measurement for model discrimination power. The optimal decision tree models were selected based on the highest AUC value. An AUC value closer to 1 indicates a model with higher discriminative power, which is recognized as a better classifier incorrectly predicting the outcome of interest. In general, models with an AUC of 0.5 suggests no discriminatory power, 0.7 to 0.8 are considered acceptable, 0.8 to 0.9 are considered excellent, and more than 0.9 is deemed to be outstanding discriminatory power (i.e., ability to diagnose patients with and without the disease or condition based on a new set of data).

***Step 3 – Predictors of ICBT Outcomes:*** Decision tree-based classifiers provide insights on different participant groups who had shown promising results following ICBT. After identifying the two most optimal models, the model-agnostic post-hoc framework SHAP was incorporated [46,47]. This framework facilitates model interpretations and identifies the most influential factors leading to successful ICBT outcomes (i.e., reduction in TFI scores following the ICBT intervention). SHAP measures the impact of variables taking into account the interaction with other variables. The SHAP values calculate the importance of a feature by comparing what a model predicts with and without the feature. However, since the order in which a model sees features can affect its predictions, this is done in every possible order so that the features are compared in a fair manner.

***Step 4 – Identification of Participants Most Likely to Benefit from ICBT:*** The CART decision tree model was used to identify the participants who are most (or least) likely to benefit from ICBT. In training, a minimum split of 20 and a max depth of 10 were used as the control parameters for the CART decision tree models. Tree pruning was utilized to reduce the overfitting in the CART decision tree models, although the best decision tree model remains the same, even after pruning.

The data analysis was performed using R (version 4.0.3) software. The code is available in the GitHub: https://github.com/Hansapani/AI-Tinnitus. Data are available on a reasonable request.

## Results

### ICBT Effects

Undertaking ICBT significantly reduced tinnitus severity ($t$(df)=16.37 (227); $P < .001$) from a mean baseline severity score of 57.93 (SD 19.17), compared with a post-ICBT severity of 34.22 (SD 22.78), as measured by the TFI. A clinically significant 13-point change in TFI score was obtained for 150 of the 228 participants (66%) post-intervention.

### Decision Tree Model Performance Evaluations

Table 1 contains the model evaluation information of all six decision tree classifiers based on the test data. Following training using the 3-fold cross-validation method, the mean accuracies of the six decision tree classifiers ranged between a minimum of 56.3% (with C5.0) to a maximum of 71.8% (with GB). Model predictions showed variations in their sensitivity (range: 68.6% to 78.3%), specificity (range: 31.1% to 64.0%), and AUC values (range: .52 to .69).

None of the six models were considered robust, as their AUC values were all below .80. As the CART and GB classifiers were found to be superior compared to the other three models when considering all evaluation measurement as a whole (accuracy, sensitivity, specificity and the AUC values), these two models were further examined.

Table 1: Decision tree model evaluations

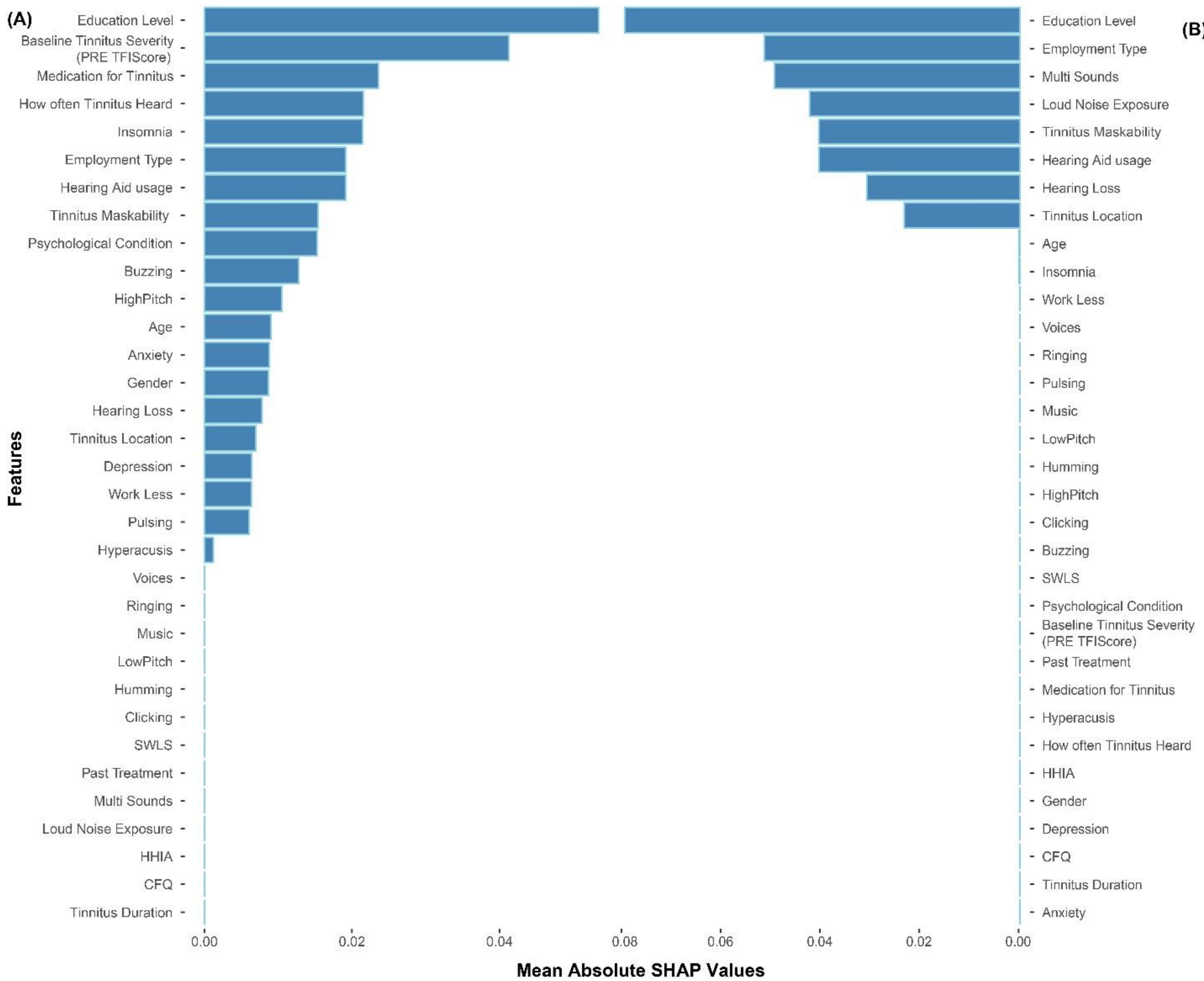
| **Classification model** | **Accuracy (%)** | **Sensitivity (%; True positive rate)** | **Specificity (%; True negative rate)** | **Area under the ROC curve (AUC)** |
|---|---|---|---|---|
| CART | 70.7±2.4 | 74.0±5.5 | 64.0±3.7 | 0.69 ± 0.001 |
| C5.0 | 56.3±1.1 | 68.6±1.9 | 31.1±6.3 | 0.52 ± 0.001 |
| Gradient Boosting (GB) | 71.8±1.5 | 78.3±2.8 | 58.7 ± 4.2 | 0.68±0.02 |
| XGBoost | 65.0±4.1 | 77.9±8.7 | 39.2±6.6 | 0.62±0.08 |
| AdaBoost algorithm | 63.6±3.2 | 73.3±5.2 | 44.0±7.8 | 0.58 ± 0.05 |
| Random Forest (RF) | 66.7±3.0 | 75.0±6.1 | 50.0±7.2 | 0.60±0.01 |

### Feature Importance

The SHAP framework was applied on the CART and GB classifiers to estimate each predictor variable's importance in predicting the ICBT outcome (see Figure 1). Larger SHAP values indicate relatively higher importance in their feature contribution. The education level (with an average SHAP value: GB = .053, CART = .079, has been identified as the most important influencing factor under both models. Although not in the same order, other features ranked within the top 10 features for both models were:, employment type (GB = .019, CART = .051), hearing aid usage (GB = .019, CART = .040), and tinnitus maskability (GB = .015, CART =

.040). Differences between these models were that the GB model ranked having the baseline tinnitus severity (GB = .041), how often tinnitus is heard (GB = .022), insomnia (GB = .021), use of medication for tinnitus(GB=.024) and a psychological condition (GB = .015), among the top 10 features, whereas the CART model ranked the presence of multiple sounds (CART = .049), loud noise exposure (CART = .042) and tinnitus location (CART = .023) as key features.

**Figure 1: Feature importance based on the mean absolute SHAP values (A) from the best GB model (B) from the best CART decision tree model. These SHAP values represent the absolute change in log odds indicating relatively higher importance with larger values.**




Figures 2 and 3 represent the effect of each feature category on the outcome variable as decided by the best GB and CART decision tree models. Feature impact on the two classes is given as two separate plots for each feature (1 indicates the effect on the successful treatment class and 0 for the other group). Positive SHAP values on each successful treatment group indicate a higher log odd of a 13-point or more tinnitus severity reduction on the TFI for that category, of the feature relative to the training set average and vice versa. Figures 2 and 3 both depict positive SHAP values for the participants who had a vocational training or Master's degree or above, higher levels of education (post-graduate degrees), and for those who are using a hearing aid in only one ear. As per the GB model, this reduction was more likely for participants with insomnia (scores of 14 or less on the ISI), presence of a psychological condition, and tinnitus described as a buzzing sound and for the participants who had their median baseline tinnitus severity scores (i.e., pre TFI) more than 55.2.

**Figure 2: Best GB model-based feature effects. Each graph represents a feature vs. corresponding SHAP value. Plots with 1 illustrate the impact of each feature on having a successful treatment outcome (13 point or more reduction on TFI score).**

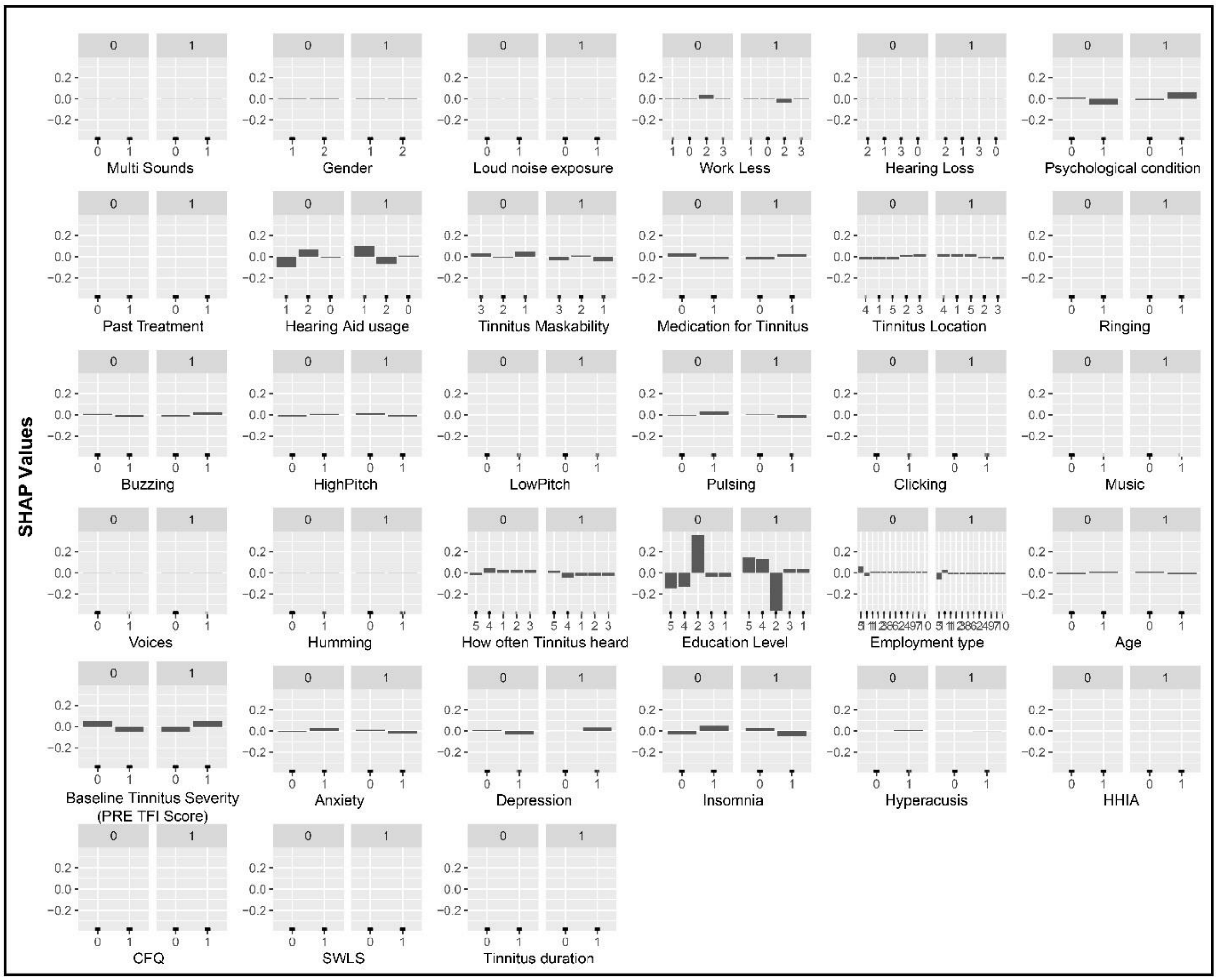


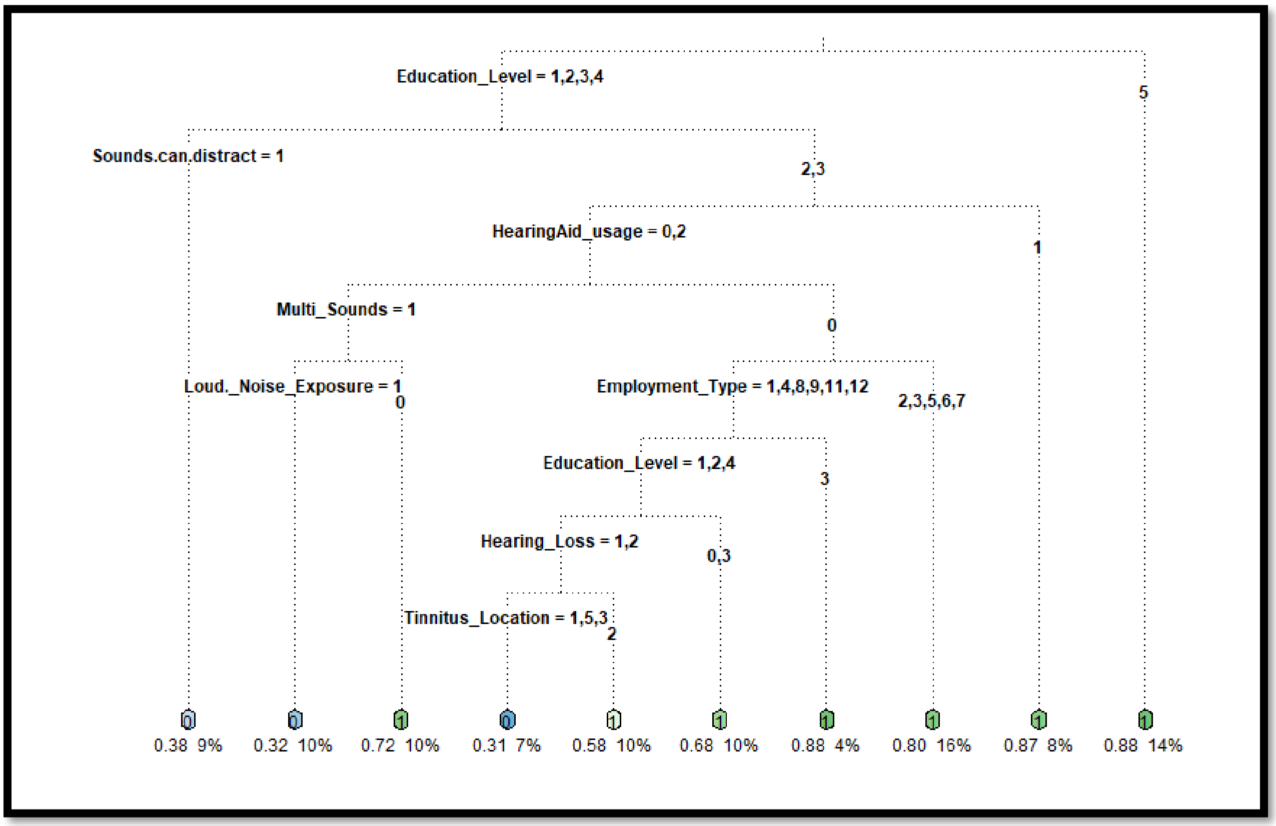
**Figure 3: Best CART decision tree model feature effects. Each graph represents a feature vs. corresponding SHAP value. Plots with 1 represent the effect of each feature on having a successful treatment outcome.**


## Identification of Participants Who are likely to Benefit from ICBT

Figure 4 presents the final decision tree model with 10 nodes. A detailed explanation is provided in Appendix 3. As we trace down from top to bottom of the decision tree, we can see how the homogeneous grouping was formed by creating binary splits at each node. The decision nodes represent the most likely treatment group (either 0 or 1) with the feature characteristics that we find as we trace down each branch of the tree. This model showed that higher education level, tinnitus maskability, hearing aid usage, the presence of multiple tinnitus sounds, loud noise exposure, employment type, the presence of hearing loss, and tinnitus location were important factors in determining treatment outcomes. The following participants' groups had shown at least an 85% chance of a TFI score reduction of 13 points or more:

- Participants with post-graduate education (master’s degree or higher).
- Participants with an education level other than master’s degree together with poor tinnitus maskability (or only partly possible) and wearing a hearing aid in one ear.
- Participants with no tinnitus maskability or only partly maskability, not wearing a hearing aid or used hearing aids bilaterally, without multiple tinnitus and having an occupation described as a professional, technical, skilled tradesman, service occupation or medical.

**Figure 4: The best CART decision tree model. The fitted tree has 10 terminal nodes (denote the decision criteria).**




## Discussion

The aim of this study was to explore predictors of outcome for ICBT for tinnitus by applying six types of decision tree models from a data set combining three clinical trials. The key findings are discussed below.

### Best Decision Tree Models

In the current study, we applied six different decision tree models to the ICBT data. Although none of the six models attain excellent or outstanding status, the CART and GB models’ discriminative power can be considered to be satisfactory, considering the moderate sample size with just 33 predictive factors (features). This is consistent with a recent study that applied 10 decision tree models on predicting the outcome of CBT in tinnitus patients (n=1,416) and found that the GB model with 26 predictive factors has the best predictive power with an AUC value of 0.89 [21].

However, further work is needed in this area to examine which models and how many vital factors may result in an optimal predictive model. A larger sample size would likely improve the results. Although, we are not sure that just adding factors would be helpful. This is because Niemann et al. [21] included 205 factors in their analysis, and of these, only 26 of them were helpful in achieving the optimal results. Moreover, of the 26 factors, only a handful had the most considerable effect. For instance, a single factor (i.e., tinnitus impairment in terms of loudness, frequency, and distress) resulted in an AUC of .79, only 3 features resulted in an AUC of .85, and only 8 factors resulted in an AUC of .85. These results indicate that including key factors with high predictive power may be a better approach than just adding all the possible factors.

**Predictors of ICBT Outcome**

Among the best decision tree models (i.e., CART and GB), various factors were found to be the critical predictors of ICBT outcome. These included demographic (i.e., education level, employment type, the presence of a psychological condition), tinnitus, and hearing-related (i.e., baseline tinnitus severity, tinnitus location, how often tinnitus is heard, having a buzzing type of tinnitus, tinnitus maskability, hearing loss type), treatment-related (i.e., hearing aid usage), and clinical factors (i.e., insomnia). However, education level was the most notable predictor among these.

Participants who had an education of master’s degree or above had an 88% chance of successful outcome. This is understandable as the ability to read, understand, and follow instructions is key for self-help interventions. However, it is likely that the way in which the materials were written may also have played a role. For instance, the UK ICBT materials were written at a 9th reading grade level [48] which may have required higher literacy skills. However, these materials have been re-written to below 6th reading grade levels [48] to ensure accessibility for those with lower education levels. More and more people, including those with lower education, are using the internet and also participate in internet-based treatments, particularly due to constraints placed on healthcare during the COVID-19 pandemic [49-51]. In the current sample, over 85% of the participants were below master level education, highlighting the need for making the ICBT more accessible to increases the chances of improved outcomes for those with lower education levels.

Baseline tinnitus severity was found to be another critical factor in predicting the ICBT outcome in the GB model. Our previous studies on the one-year outcome [52], and also application of the univariate and multivariate analysis on the current sample [19], have identified baseline tinnitus severity as a critical predictive variable. In the Niemann et al. [21] study, tinnitus loudness, frequency, and distress measured using the visual analog scale was found to be the kye predictive factor, although the tinnitus distress measured using the German version of the Tinnitus Questionnaire (TQ), which is comparable to TFI in this study was not found to be the key predictive factor. However, both clinical experiences as well findings from many previous studies have identified baseline tinnitus severity to be an important factor in determining treatment outcomes. The clinical factors; depression and anxiety were among key predictive factors in the GB model. A recent clinical trial by Beukes et al. [52] as well as the Niemann et al. [21] study identified that those with high depression had a better chance of success.

Although various other tinnitus and hearing-related variables could play a role in determining the outcome of ICBT, predictive power is relatively low based on the SHAP values. Nevertheless, it is useful for hearing healthcare professionals examine these factors when deciding the candidacy for self-help psychological interventions such as the ICBT. Moreover, it would be useful for future studies to examine any additional factors (e.g., health literacy) that may have a bearing on the ICBT outcome.

**Study Limitations and Future Directions**

Although the study is among the first to apply the data mining models to the ICBT data, it has several limitations. The sample size was limited, which may have contributed to the low predictive accuracies of the models. The exploratory decision tree models worked better when including a large number of predictive factors. In the current study, we included 33 predictive factors, which may be limited and may have missed some important factors (e.g., health literacy) that have a bearing on ICBT outcome.

Moreover, the inclusion/exclusion criteria used in the three trials from which this data was generated may have resulted in a sample with higher tinnitus severity, which may not represent the general tinnitus population, which may have further contributed to limited key findings. Future studies could include more extensive samples of heterogeneous tinnitus patients as well as include all the possible predictive factors which could help with improving the predictive power. Moreover, developing nonlinear classifiers with artificial neural networks (ANN), and support vector machine (SVM) could help in achieving higher prediction accuracies and should be examined in future studies.

In conclusion, tree models, especially the CART and GB models, appear to be promising in predicting the ICBT outcomes. Future studies should be undertaken with larger sample sizes and including a more comprehensive range of predictive factors to improve their predictive power.

## Conflicts of Interest

None declared.

## References

1. McCormack A, Edmondson-Jones M, Somerset S, Hall D. A systematic review of the reporting of tinnitus prevalence and severity. Hearing research 2016 Jul; 337: 70-79. [doi: 10.1016/j.heares.2016.05.009]

2. Kostev K, Alymova S, Kössl M, Jacob L. Risk Factors for Tinnitus in 37,692 Patients Followed in General Practices in Germany. Otol neurotol. 2019 Apr; *40*(4), 436–440. [doi:10.1097/MAO.0000000000002161]

3. Kim HJ, Lee HJ, An SY, Sim S, Park B, Kim SW, et al. Analysis of the prevalence and associated risk factors of tinnitus in adults. PloS one 2015 May 28; 10(5): e0127578. [doi: 10.1371/journal.pone.0127578]

4. Beukes EW, Manchaiah V, Allen PM, Andersson G, Baguley D. Exploring tinnitus heterogeneity. Prog Brain Res. 2021; 260:79-99. [doi: 10.1016/bs.pbr.2020.05.022]

5. Manchaiah V, Beukes E, Granberg S, Durisala N, Baguley D, Allen P, et al. Problems and life effects experienced by tinnitus research study volunteers: An Exploratory study using the ICF classification. Journal of the American Academy of Audiology 2018 Nov/Dec; 29(10), 936-947. [doi:10.3766/jaaa.17094]

6. Beukes EW, Manchaiah V, Andersson G, Allen PM, Terlizzi, PM, Baguley DM. Situationally influenced tinnitus coping strategies: a mixed methods approach. Disability and rehabilitation 2018 Dec; 40(24): 2884–2894. [doi: 10.1080/09638288.2017.1362708]

7. Salazar JW, Meisel K, Smith ER, Quiggle A, McCoy DB, Amans MR Depression in Patients with Tinnitus: A Systematic Review. Otolaryngology head and neck surgery 2019 Jul;161(1): 28–35. [doi: 10.1177/0194599819835178]

8. Trevis KJ, McLachlan NM, Wilson SJ. A systematic review and meta-analysis of psychological functioning in chronic tinnitus. Clinical psychology review 2018 Mar; 60: 62–86. [doi: 10.1016/j.cpr.2017.12.006]

9. Fuller T, Cima R, Langguth B, Mazurek B, Vlaeyen JWS, Hoare DJ. Cognitive behavioural therapy for tinnitus. Cochrane Database of Systematic Reviews 2020 Jan; 1: CD012614. [doi: 10.1002/14651858.CD012614.pub2]

10. Landry EC, Sandoval X, Simeone CN, Tidball G, Lea J, Westerberg BD. Systematic Review and Network Meta-analysis of Cognitive and/or Behavioral Therapies (CBT) for Tinnitus Otology & neurotology 2020 Feb; 41(2): 153–166. [doi: 10.1097/MAO.0000000000002472]

11. Tunkel D, Bauer C, Sun G, Rosenfeld R, Chandrasekhar S, Cunningham Jr E, et al. Clinical practice guideline: Tinnitus. Otolaryngology–Head and Neck Surgery 2014 Oct; 151(4): 533-541. [doi: 10.1177/0194599814547475].

12. Andersson G, Strömgren T, Ström L, Lyttkens L. Randomized controlled trial of internet-based cognitive behavior therapy for distress associated with tinnitus. Psychosomatic medicine 2002 Sep-Oct; 64(5): 810–816. [doi: 10.1097/01.psy.0000031577.42041.f8]

13. Beukes EW, Vlaescu G, Manchaiah V, Baguley DM, Allen, PM, Kaldo V. et al. Development and technical functionality of an Internet-based intervention for tinnitus in the UK. Internet Interventions 2016 Nov; 6: 6-15. [doi:10.1016/j.invent.2016.08.002].

14. Beukes EW, Manchaiah V, Allen PM, Baguley DM, Andersson G. Internet-Based Interventions for Adults With Hearing Loss, Tinnitus, and Vestibular Disorders: A Systematic

Review and Meta-Analysis. Trends in hearing 2019 Jan; 23: 2331216519851749. [doi: org/10.1177/2331216519851749]

15. Beukes EW, Andersson G, Fagelson M, Manchaiah V. Audiologist-supported Internet-based Cognitive Behavioral Therapy for Tinnitus in the United States: A Pilot Trial. American Journal of Audiology (In Press).

16. Beukes, E.W., Andersson, G., Fagelson, M.A., & **Manchaiah, V.** (Submitted). Internet-based audiologist-guided Internet-based cognitive behavioral therapy for tinnitus in the United States: A randomized controlled trial. Journal of Medical Internet Research. https://preprints.jmir.org/preprint/27584

17. Beukes EW, Baguley DM, Allen PM, Manchaiah V, Andersson G. Audiologist-guided Internet-based cognitive behavior therapy for adults with tinnitus in the United Kingdom: A randomized controlled trial. Ear and Hearing 2018 May-Jun; 39(3): 423–433. [doi: 10.1097/AUD.0000000000000505]

18. Beukes EW, Andersson G, Allen PM, Manchaiah V, Baguley DM. Effectiveness of guided Internet-based Cognitive Behavioral Therapy vs face-to-face clinical care for treatment of tinnitus: A randomized clinical trial. JAMA Otolaryngology—Head & Neck Surgery 2018 Dec; 144(12): 1126–1133. [doi.:10.1001/jamaoto.2018.2238]

19. Rodrigo HS, Beukes EW, Andersson G, Manchaiah V. Predictors of outcomes of internet-based cognitive behavioural therapy intervention for individuals with tinnitus. BMJ Open 2021;, 11, e049384. [doi: 10.1136/bmjopen-2021-049384]

20. Juhola M, Viikki K, Laurikkala J, Aurama Y, Kentala E, Pyykko I. Application of artificial intelligence in audiology. Scand. Audiol. Suppl. 2001; 52:97 - 99. [doi: 10.1080/010503901300007209]

21. Niemann U, Boecking B, Brueggemann P, Mebus W, Mazurek B, Spiliopoulou, M. Tinnitus-related distress after multimodal treatment can be characterized using a key subset of baseline variables. PloS One 2020 Jan 30; 15(1): e0228037. [doi.:10.1371/journal.pone.0228037]

22. Rasku J, Pyykkö I, Levo H, Kentala E, Manchaiah V. Disease Profiling for Computerized Peer Support of Ménière's Disease. JMIR rehabilitation and assistive technologies 2015 Sep 3; *2*(2): e9. [doi:10.2196/rehab.4109]

23. Shoushtarian M, Alizadehsani R, Khosravi A, Acevedo N, McKay CM, Nahavandi S, et al. Objective measurement of tinnitus using functional near-infrared spectroscopy and machine learning. PloS one 2020 Nov 18; 15(11): e0241695. [doi:10.1371/journal.pone.0241695]

24. Weininger O, Warnecke A, Lesinski-Schiedat A, Lenarz T, Stolle S. Computational analysis based on audioprofiles: A new possibility for patient stratification in office-based otology. Audiology research 2019 Nov 5; 9(2): 230. [doi:10.4081/audiores.2019.230]

25. Raj, S. (2019, August 9). Effects of Multi-collinearity in Logistic Regression, SVM, Random Forest (RF). Retrieved from https://medium.com/@raj5287/effects-of-multi-collinearity-in-logistic-regression-svm-rf-af6766d91f1b

26. Beukes EW, Allen PM, Manchaiah V, Baguley DM, Andersson G. Internet-Based intervention for tinnitus: outcome of a single-group open trial. J Am Acad Audiol 2017 Apr; 28(4):340–351. [doi: 10.3766/jaaa.16055].

27. Beukes E, Andersson G, Manchaiah V, Kaldo V. Cognitive behavioral therapy for tinnitus. USA: Plural Publishing Inc; 2021.

28. Manchaiah V, Vlaescu G, Varadaraj S, Aronson EP, Fagelson MA, Munoz MF, et al. Features, Functionality, and Acceptability of Internet-Based Cognitive Behavioral Therapy for Tinnitus in the United States. American Journal of Audiology 2020 Sep 3; 29(3): 476–490. [doi: 10.1044/2020_AJA-20-00002]

29. Vlaescu G, Alasjö A, Miloff A, Carlbring P, Andersson G. Features and functionality of the Iterapi platform for Internet-based psychological treatment. Internet Interventions 2016 Nov; 6: 107–114. [doi: 10.1016/j.invent.2016.09.006]

30. Meikle MB, Henry JA, Griest SE, Stewart BJ, Abrams HB, McArdle Rachel, et al. The tinnitus functional index: development of a new clinical measure for chronic, intrusive tinnitus. Ear Hear 2012 Mar-Apr; 33(2):153–176. [doi: 10.1097/AUD.0b013e31822f67c0].

31. Bastien CH, Vallieres A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001 Jul; 2(4): 297–307. [doi: 10.1016/s1389-9457(00)00065-4].

32. Spitzer RL, Kroenke K, Williams JB, Lowe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22; 166(10); 1092–1097. [doi: 10.1001/archinte.166.10.1092].

33. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999 Nov 10; 282(18): 1737–1744. [10.1001/jama.282.18.1737].

34. Newman CW, Weinstein BE, Jacobson GP, Hug GA. Test-retest reliability of the hearing handicap inventory for adults. Ear Hear 1991 Oct; 12(5): 355–357. [doi: 10.1097/00003446-199110000-00009].

35. Khalfa S, Dubal S, Veuillet E, et al. Psychometric normalization of a hyperacusis questionnaire. ORL J Otorhinolaryngol Relat Spec 2002 Nov-Dec; 64(6): 436–442. [doi: 10.1159/000067570].

36. Broadbent DE, Cooper PF, FitzGerald P, Parkes KR. The Cognitive Failures Questionnaire (CFQ) and its correlates. Br J Clin Psychol 1982 Feb;21(1):1–16. [doi: 10.1111/j.2044-8260.1982.tb01421.x.]

37. Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction With Life Scale. J Pers Assess 1985 Feb; 49(1): 71–75. [doi: 10.1207/s15327752jpa4901_13]

38. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning: with applications in R. USA: Springer; 2013.

39. Breiman, L. Random forests. Machine learning 2001 Oct; 45: 5–32.

40. Friedman JH. Greedy function approximation: A gradient boosting machine. Annals of Statistics 2001 Oct;29(5),1189–1232. [doi: 10.1214/aos/1013203451]

41. Chen, T, Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM; 2016 Aug 13;785–794. New York, NY, USA. [doi:10.1145/2939672.2939785]

42. Breiman L, Friedman J, Sone CJ, Olshen, RA. Classification and Regression Trees. USA; Chapman and Hall. 1984.

43. Quinlan JR. C4.5: Programs for Machine Learning. USA; Morgan Kaufmann Publishers 1993.

44. Gandhi R. Boosting Algorithms: AdaBoost, Gradient Boosting and XGBoost. 2018 May 06; Retrieved from https://hackernoon.com/gradient-boosting-and-xgboost-90862daa6c77.

45. Gini, C. On the Measure of Concentration with Special Reference to Income and Statistics. Colorado College Publication 1936; 208:73–79.

46. Lundberg SM, Lee Su-In. A Unified Approach to Interpreting Model Predictions. NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017 Dec 04;4768-4777.

47. Lundberg SM, Erion G, Chen, H, DeGrave A, Prutkin JM, Nair Bala et. al. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence 2020 Jan 17;2:56-67.

48. Beukes EW, Fagelson M, Aronson EP, Munoz MF, Andersson G, Manchaiah V. Readability following cultural and linguistic adaptations of an Internet-based intervention for tinnitus for use in the United States. American Journal of Audiology 2020 Jun 08; 29(2):97–109. [doi. 10.1044/2019_AJA-19-00014]

49. Beukes EW, Baguley DM, Jacquemin L, Lourenco M, Allen PM, Onozuka J, et al. Changes in Tinnitus Experiences During the COVID-19 Pandemic. Frontiers in public health 2020 Nov 05;8:592878. [doi: 10.3389/fpubh.2020.592878]

50. Beukes EW, Onozuka J, Brazell T, Manchaiah, V. Coping with tinnitus during the COVID-19 pandemic. American Journal of Audiology (In press)

51. Wind TR, Rijkeboer M, Andersson G, Riper H. The COVID-19 pandemic: The 'black swan'for mental health care and a turning point for e-health. Internet interventions 2020 Mar;20. [doi:10.1016/j.invent.2020.100317]

52. Beukes EW, Manchaiah V, Baguley DM, Allen PM, Andersson G. Long-term efficacy of Audiologist-guided Internet-based Cognitive Behaviour Therapy for Tinnitus. American Journal of Audiology 2018 Nov 19; 27(3S):431-447. [doi: 10.1044/2018_AJA-IMIA3-18-0004]

53. Beukes, E.W., Andersson, G., Fagelson, M.A., & Manchaiah, V. (2021). Dismantling internet-based cognitive behavioral therapy for tinnitus. The contribution of applied relaxation: A randomized controlled trial. Internet Interventions, 25, 100402. https://doi.org/10.1016/j.invent.2021.100402

## Abbreviations

AI: Artificial Intelligence
AUC: Area Under the receiver operation characteristic Curve
CART : Classification and Regression Trees
CBT : Cognitive Behavioral Therapy
CFQ : Cognitive Failures Questionnaire
GAD: Generalized Anxiety Disorder
GB: Gradient Boosting
HHIA: Hearing Handicap Inventory for Adults Screening
HQ: Hyperacusis Questionnaire
ICBT: Internet-based Cognitive Behavioral Therapy
ISI: Insomnia Severity Index
ML: Machine Learning
PHQ: Patient Health Questionnaire
RF: Random Forest
ROC: Receiver Operating Characteristic Curve
SHAP: SHapley Additive exPlanations
SWLS: Satisfaction with Life Scales
TFI: Tinnitus Functional Index
XGBoost : eXtreme Gradient Boosting

## Appendix 1: Participants characteristics

Table 1.1: Characteristic of the study participants

| Characteristic | N (%) | Mean (SD) |
|---|---|---|
| **Demographic characteristics** | | |
| Age (in years) | | 55.14 (12.92) |
| Gender | | |
| ▪ Female | 98 (43%) | |
| ▪ Male | 130 (57%) | |
| Highest level of education | | |
| ▪ High school or below | 59 (26%) | |
| ▪ College | 47 (21%) | |
| ▪ Vocational training | 31 (13%) | |
| ▪ Bachelor's degree | 61 (26%) | |
| ▪ Masters degree or above | 30 (13%) | |
| Employment | | |
| ▪ Manager | 27 (12%) | |
| ▪ Professional | 46 (20%) | |
| ▪ Technical | 16 (6%) | |
| ▪ Administrative | 17 (7%) | |
| ▪ Skilled tradesman | 11 (5%) | |
| ▪ Service occupation | 11 (5%) | |
| ▪ Medical | 6 (3%) | |
| ▪ Sales | 8 (3%) | |
| ▪ Homemaker | 4 (2%) | |
| ▪ Student | 1 (0%) | |
| ▪ Retired | 73 (32%) | |
| ▪ Unemployed | 11 (5%) | |
| Loud noise exposure | | |
| ▪ Yes | 103 (45%) | |
| ▪ No | 125 (55%) | |
| Diagnosed with a psychological condition | | |
| ▪ Yes | 50 (22%) | |
| ▪ No | 178 (78%) | |
| Working less due to tinnitus | | |
| ▪ Reduced hours | 8 (4%) | |
| ▪ Stopped work | 32 (14%) | |
| ▪ Disability allowance | 7 (3%) | |
| ▪ No | 181 (79%) | |
| **Tinnitus and hearing-related characteristics** | | |
| Baseline tinnitus severity (aka: Pre-TFI, measured using Tinnitus Functional Index) | | 57.93 (19.17) |
| Tinnitus duration (in years) | | 17.68 (19.42) |
| How often tinnitus is heard | | |
| ▪ Occasionally | 4 (2%) | |
| ▪ When taking out my hearing aid(s) | 3 (1%) | |
| ▪ At night | 4 (2%) | |
| ▪ Most of the time | 63 (27%) | |
| ▪ All the time | 154 (68%) | |

| | | |
|---|---|---|
| Tinnitus location | | |
| ▪ One ear | 61 (27%) | |
| ▪ Both ears | 109 (48%) | |
| ▪ In my head | 34 (15%) | |
| ▪ Other location | 3 (1%) | |
| ▪ Unsure | 21 (9%) | |
| Type of tinnitus sound (answering Yes) | | |
| ▪ Ringing | 71 (31%) | |
| ▪ Buzzing | 75 (33%) | |
| ▪ High pitched sound | 130 (57%) | |
| ▪ Low pitched sound | 16 (7%) | |
| ▪ Pulsating | 28 (12%) | |
| ▪ Clicking | 14 (6%) | |
| ▪ Music | 4 (2%) | |
| ▪ Voices | 3 (1%) | |
| ▪ Humming | 21 (9%) | |
| Multiple sounds heard | | |
| ▪ Yes | 73 (32%) | |
| ▪ No | 155 (68%) | |
| Presence of a hearing loss | | |
| ▪ No | 49 (21%) | |
| ▪ Both ears | 104 (46%) | |
| ▪ One ear | 46 (20%) | |
| ▪ Unsure | 29 (13%) | |
| **Treatment-related characteristics** | | |
| Past tinnitus treatment sought | | |
| ▪ Yes | 58 (25%) | |
| ▪ No | 170 (75%) | |
| Sounds can distract from tinnitus (tinnitus | | |
| maskability) | 26 (11%) | |
| ▪ Fully | 178 (78%) | |
| ▪ Partially | 24 (10%) | |
| ▪ Not at all | | |
| Hearing aid use | | |
| ▪ No | 159 (70%) | |
| ▪ Unilateral | 19 (8%) | |
| ▪ Bilateral | 50 (22%) | |
| Medication use | | |
| ▪ Yes | 130 (57%) | |
| ▪ No | 98 (43%) | |

## Appendix 2: Predictor Variables

Table 2.1: Demographic variables (7 variables)

| **Variable** | **Question** | **Response options** |
|---|---|---|
| Age | What is your age? | In years<br><br>Split into dichotomous variables (<=57 years of age and >57 years of age) based on the median |
| Gender | What is your gender? | Male (1), Female (2) |
| Education level | What is the highest level of education you have completed? | Highschool or less (1), College (2), Vocational training (3), Bachelor’s degree (4), Master’s degree or above (5) |
| Employment type | What best describes your employment? | Manager (1), Professional (2), Technical (3), Administrative (4), Skilled tradesman (5), Service occupation (6), Medical (7), Sales (8), Home maker (9), Student (10), Retired (11), Unemployed (12) |
| Loud noise exposure | Have you been exposed to loud noise? | Yes (1) , No (0) |
| Diagnosed with psychological condition | Have you been presently diagnosed with any psychological conditions including anxiety and depression? | Yes (1) , No (0) |
| Work less due to tinnitus | Do you work less because of your tinnitus? | No (0), Reduced hours (1), Stopped work (2), Disability allowance (3) |

Table 2.2: Tinnitus and hearing-related variables (15 variables)

| **Variable** | **Question** | **Response options** |
|---|---|---|
| Baseline tinnitus severity (Pre-TFI) | Measured using the Tinnitus Functional Index (TFI) | Scores range from 0 to 100.<br><br>Split into dichotomous variables (<=55.2 and >55.2) based on the median |

| | | |
|---|---|---|
| Tinnitus duration | How long have you had tinnitus for? | In years<br><br><br>Split into dichotomous variables (<=10.00 years and >10.00 years) based on the median |
| How often is tinnitus heard? | How often is tinnitus heard? | Occasionally (1), When taking out my hearing aid(s) (2), At night (3), Most of the time (4), All the time (5) |
| Tinnitus location | Where do you notice your tinnitus? | One ear (1), Both ears (2), In my head (3), Unsure (4), Other (5) |
| Type of tinnitus (9 different types) | ▪ Ringing<br>▪ Buzzing<br>▪ High pitched sound<br>▪ Low pitched sound<br>▪ Pulsing<br>▪ Clicking<br>▪ Music<br>▪ Voices<br>▪ Humming | For each item: Yes (1) , No (0) |
| Multiple tones heard | This variable is computed based on responses to types of tinnitus. Answer yes to multiple types of tinnitus was considered as multiple tones heard | Yes (1) , No (0) |
| Presence of a hearing loss | Do you have a hearing loss? | No (0), Both ears (1), One ear (2), Unsure (3) |

Table 2.3: Treatment-related variables (4 variables)

| **Variable** | **Question** | **Response options** |
|---|---|---|
| Past treatment sought | Have you received treatment for tinnitus in the past? | Yes (1) , No (0) |
| Sounds can distract from tinnitus (tinnitus maskability) | How well can sounds around you distract you from your tinnitus or make the tinnitus less noticeable? | Fully (1), Partially (2), Not at all (3) |
| Hearing aid use | Do you wear hearing aid(s) or any other amplification devices? | No (0), One ear (1), Both ears (2) |
| Medication use | Do you currently take any medications? | Yes (1) , No (0) |

Table 2.4: Clinical factors (7 variables)

| **Variable** | **Questionnaire** | **Number of items/ Response options** | **Score** |
|---|---|---|---|
| Anxiety | General Anxiety Disorders (GAD-7) | 7-items<br><br>4-point scale with "not at all" (score of 0) to "nearly every day" (score of 3) | Higher number indicates more severe anxiety (scores range between 0–21). The total score is interpreted as follows:<br>▪ 0–4: minimal anxiety<br>▪ 5–9: mild anxiety<br>▪ 10–14: moderate anxiety<br>▪ 15–21: severe anxiety<br><br>Split into dichotomous variables (<=9 no anxiety and >9 anxiety) |
| Depression | Patient Health Questionnaire (PHQ-9) | 9-items<br><br>4-point scale with "not at all" (score of 0) to "nearly every day" (score of 3) | Higher number indicates more severe depression (scores range between 0–27).<br><br>The total score is interpreted as follows:<br>▪ 5–9: mild depression<br>▪ 10–14: moderate<br>▪ 15–19: moderately severe<br>▪ 20–18: severe depression<br><br>Split into dichotomous variables (<=14 no depression and >14 depression) |
| Insomnia | Insomnia Severity Index (ISA) | 7-item<br><br>5-point scale with "no problem" (score of 0) to "very severe problem" (score of 4) | Higher number indicates more severe insomnia (scores range between 0–28).<br><br>The total score is interpreted as follows:<br>▪ 0–7: not clinically significant<br>▪ 8–14: subthreshold insomnia<br>▪ 15–21: clinical insomnia (moderate severity)<br>▪ 22–28: clinical insomnia (severe degree)<br><br>Split into dichotomous variables (<=14 no insomnia and >15 insomnia) |
| Hyperacusis | Hyperacusis Questionnaire (HQ) | 14-items<br><br>4-point scale with "no" (score of 0) to | Higher number more severe hyperacusis (scores range between 0–42). |

| | | | |
|---|---|---|---|
| | | "yes, a lot" (score of 3) | The total score is interpreted as follows:<br>▪ >28: strong hypersensitivity<br><br>Split into dichotomous variables (<=28 no hyperacusis and >28 hyperacusis) |
| Hearing disability | Hearing Handicap Inventory for Adults – Screening (HHIA-S) | 10-items<br><br>3-point scale with "yes" (score of 4) to "no" day (0) | Higher number more severe hearing disability (scores range between 0–40).<br><br>The total score is interpreted as follows:<br>▪ 0–8: no hearing disability<br>▪ 10–24: mild to moderate hearing disability<br>▪ 26–40: severe hearing disability<br><br>Split into dichotomous variables (<=8 no hearing disability and >=10 hearing disability) |
| Cognitive failures | Cognitive Failures Questionnaire (CFQ) | 25-items<br><br>5-point scale with "never" (score of 0) to "very often" (score of 4) | Higher scores indicate more difficulties (cognitive failures) in perception, memory, and motor function (score range 0–100).<br><br>The total score is interpreted as follows:<br>The scores range 0–100 with higher scores indicating more cognitive failures/problems (or reduced cognitive functioning).<br><br>Split into dichotomous variables (<=32 no cognitive problems and >32 cognitive problems) |
| Life satisfaction | Satisfaction with Life Scale (SWLS) | 5-items<br><br>7-point scale with "strongly disagree" (score of 1) to "strongly agree" (7) | Higher number indicated more satisfaction with life (scores range between 5–35).<br><br>The total score is interpreted as follows:<br>▪ 0–9: extremely dissatisfied<br>▪ 10–14: dissatisfied<br>▪ 15–19: below average satisfaction |

| | | | |
|---|---|---|---|
| | | | ▪ 20–24: average satisfaction<br>▪ 25–29: high satisfaction<br>▪ 30–35: highly satisfied<br><br>Split into dichotomous variables (<=19 life satisfaction and >19 high satisfaction) |

## Appendix 3: Detailed explanation of the best CART decision tree model

Figure 4 in the manuscript presents the final CART decision tree model which showed that education level, sounds can distract, hearing aid usage, multi-sounds, loud noise exposure, employment type, hearing loss, and tinnitus location to be important factors. In fact, SHAP analysis discussed above had identified these factors to be the most influential features contributing to the model outcome. In the following, we summarize the characteristics of the participant groups who were most likely to be benefitted from the ICBT intervention.

- If a participant has an education of master's degree or above, he/she has approximately 88% chance of having successful outcome with the ICBT intervention. 14% of the study participants were in this category.
- If a participant has a different education level other than master's, and if she/he is not distracted at all or has a partial distraction by the sounds from the tinnitus, and if he/she uses hearing aid in his/her single ear, he/she has about 87% chance of success. 8% of the study participants were in this category.
- If a participant has a different education level other than master's, and if she/he is not distracted at all or has a partial distraction by the sounds from the tinnitus, and if he/she uses no hearing aid or hearing aids for both ears, but no multiple tones were heard and works in one of the following; professional, technical, skilled tradesman, service occupation or medical, then he/she has about 80% chance of success. 16% of the study participants were in this category.
- If a participant is not distracted at all or has a partial distraction by the sounds from the tinnitus, and if he/she uses no hearing aid or hearing aids for both ears, but no multiple tones were heard and works in one of the following; manager, administrative, sales, home maker, student, retired or unemployed whom had a vocational training, then he/she has about 88% chance of success. Only 4% of the study participants were in this category.
- If a participant is not distracted at all or has a partial distraction by the sounds from the tinnitus, and if he/she uses no hearing aid or hearing aids for both ears, but no multiple tones were heard and works in one of the following; manager, administrative, sales, home maker, student, retired or unemployed whom had an education either high school or less, college or bachelor's degree, and he/she has no hearing loss or if it is unsure, then he/she has about 68% chance of success. 10% of the study participants were in this category.
- If a participant is not distracted at all or has a partial distraction by the sounds from the tinnitus, and if he/she uses no hearing aid or hearing aids for both ears, but no multiple

tones were heard and works in one of the following; manager, administrative, sales, home maker, student, retired or unemployed whom had an education either high school or less, college or bachelor's degree, and he/she has a hearing loss either in one ear or both, and he/she recognized his/her tinnitus location to be in their head, then he/she has about 58% chance of success. 10% of the study in this category.

- If a participant has a different education level other than master's and he/she is not distracted at all or has a partial distraction by the sounds from the tinnitus, and if he/she uses no hearing aid or hearing aids for both ears, and he/she hears multiple tones and have not had a loud noise exposure, then he/she has about 72% chance of success. 10% of the study participants were in this category.

Next, we summarize the characteristics of the subject groups whom least likely to be benefitted with the ICBT treatment.

- If a participant has a different education level other than masters', and if she/he is fully distracted by the sounds, then he/she has about 38% chance of success. 9% of the study participants were in this category.
- If a subject has a different education level other than master's and he/she is not distracted at all or has a partial distraction by the sounds from the tinnitus, and if he/she uses no hearing aid or hearing aids for both ears, and he/she hears multiple tones and had a loud noise exposure, then he/she has about, only 32% chance of success. 10% of the study participants were in this category.
- If a subject is not distracted at all or has a partial distraction by the sounds from the tinnitus, and if he/she uses no hearing aid or hearing aids for both ears, but no multiple tones were heard and works in one of the following; manager, administrative, sales, home maker, student, retired or unemployed whom had an education either high school or less, college or Bachelor's degree, and he/she has a hearing loss either in one ear or both, and he/she recognized his/her tinnitus location to be in somewhere other than the head, then he/she has about 31% chance of success. Only 7% of the study participants were in this category.

These subject groupings had shown consistent probabilities in majority of the CART decision tree models that were generated with different random initializations.