Predicting hearing help-seeking: What features are important for a profiling module within a hearing mHealth application?

Giulia Angonese, Mareike Buhl, Inka Kuhlmann, Birger Kollmeier, Andrea Hildebrandt

Submitted to: JMIR Human Factors
on: August 30, 2023

Table of Contents

Predicting hearing help-seeking: What features are important for a profiling module within a hearing mHealth application?

Giulia Angonese[1, 2]; Mareike Buhl[2, 3]; Inka Kuhlmann[1, 2]; Birger Kollmeier[2, 4, 5]; Andrea Hildebrandt[1, 2, 5]

[1]Department of Psychology Psychological Methods and Statistics Carl von Ossietzky Universität Oldenburg Oldenburg DE
[2]Cluster of Excellence 'Hearing4all' Oldenburg DE
[3]Institut Pasteur Université Paris Cité Inserm, Institut de l'Audition Paris FR
[4]Department of Medical Physics and Acoustics Carl von Ossietzky Universität Oldenburg Oldenburg DE
[5]Research Center Neurosensory Science Carl von Ossietzky Universität Oldenburg Oldenburg DE

Corresponding Author:
Giulia Angonese
Department of Psychology
Psychological Methods and Statistics
Carl von Ossietzky Universität Oldenburg
Ammerländer Heerstraße 114-118
Oldenburg
DE

Abstract

Background: Mobile health care solutions can improve quality, accessibility and equity of health services, fostering early rehabilitation. For people suffering from hearing loss, mobile applications might be designed to support the decision-making processes in auditory diagnostics and to provide treatment recommendations to the user (e.g., hearing aid need). For some individuals, such mobile app might be the first contact with a hearing diagnostic service and should motivate users with hearing loss to seek professional help.

Objective: This study aims at characterizing individuals who are more or less prone to seek professional help after the repeated use of an app-based hearing test. The goal is to develop a profiling module building upon hearing related traits and personality characteristics to secure personalized treatment recommendations in hearing mHealth solutions.

Methods: N=185 (106 females) non-aided older individuals (Mage=63.8, SDage=6.6) with subjective hearing loss participated in a comprehensive online study. We collected cross-sectional and longitudinal data on several hearing-related and psychological features that were previously found to predict hearing help-seeking. Readiness to seek help was assessed as outcome variable at study-end and after two months. Participants were classified into help-seekers and non-seekers with several supervised machine learning algorithms (Random Forest, Naïve Bayes and Support Vector Machine). The most relevant features for prediction were identified with feature importance analysis.

Results: The algorithms correctly predicted action to seek help at study-end in 66 to 70% of cases, reaching 75% classification accuracy at follow-up. Among the most important features for classifications were the degree of hearing loss and its perceived consequences in daily life, attitude towards hearing aids, physical health and sensory-sensitivity personality.

Conclusions: This study contributes to the identification of individual characteristics that predict help-seeking in older individuals with self-perceived hearing loss. Suggestions for the implementation of an individual profiling algorithm and for targeted recommendations in hearing mHealth applications are derived.

(JMIR Preprints 30/08/2023:52310)
DOI: https://doi.org/10.2196/preprints.52310

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Predicting hearing help-seeking: What features are important for a profiling module within a hearing mHealth application?

Abstract

Background: Mobile health (mHealth) solutions can improve quality, accessibility and equity of health services, fostering early rehabilitation. For individuals suffering from hearing loss, mHealth applications might be designed to support the decision-making processes in auditory diagnostics and to provide treatment recommendations to the user (e.g., hearing aid need). For some individuals, such an mHealth app might be the first contact with a hearing diagnostic service and should motivate users with hearing loss to seek professional help in a targeted manner. But personalizing treatment recommendations is only possible by knowing the individual's profile regarding the outcome of interest.

Objective: This study aims at characterizing individuals who are more or less prone to seek professional help after the repeated use of an app-based hearing test. The goal is to derive relevant hearing related traits and personality characteristics toward personalized treatment recommendations for users of hearing mHealth solutions.

Methods: N=185 (106 females) non-aided older individuals (M_{age}=63.8, SD_{age}=6.6) with subjective hearing loss participated in a mobile study. We collected cross-sectional and longitudinal data on a comprehensive set of 83 hearing-related and psychological measures, among those previously found to predict hearing help-seeking. Readiness to seek help was assessed as outcome variable at study-end and after two months. Participants were classified into help-seekers and non-seekers with several supervised machine learning algorithms (Random Forest, Naïve Bayes and Support Vector Machine). The most relevant features for prediction were identified with feature importance analysis.

Results: The algorithms correctly predicted action to seek help at study-end in 66% to 70% of cases, reaching 75% classification accuracy at follow-up. Among the most important features for classifications, beyond hearing performance, were the perceived consequences of hearing loss in daily life, attitude towards hearing aids, motivation to seek help, physical health, sensory-sensitivity personality trait, neuroticism and income.

Conclusions: This study contributes to the identification of individual characteristics that predict help-seeking in older individuals with self-perceived hearing loss. Suggestions are made for their implementation in an individual profiling algorithm and for deriving targeted recommendations in hearing mHealth applications.

Keywords: Hearing loss; mHealth; older adults; help-seeking; mobile study; machine learning; supervised classification; feature importance; profiling.

Introduction

Mobile health solutions for hearing care

Hearing enables individuals to experience their surroundings and to communicate with others. Thus, hearing difficulties can have a strong impact on individuals' quality of life. Hearing loss (HL) is one of the most common chronic diseases worldwide and it affects 20.3% of the world's population. More than 60% of individuals with HL are older than 50 years of age, with the principal cause being age-related HL [1]. Untreated hearing difficulties have been associated with lower self-rated health, depression, and anxiety, in addition to physical and cognitive decline, dementia, and hospitalization in the older population [2–4]. The primary rehabilitative strategy for individuals with moderate to severe HL is the use of hearing aids (HA), which increases activity levels, general health, quality of life [5] and decreases social isolation and depressive symptoms [6] by supporting hearing ability and communication efficacy. Despite the positive effects of hearing rehabilitation, the prevalence of HA use is still limited to about 25% of the hearing impaired population [2,4,7,8]. Moreover, there is an average delay of nine years between the time a person first acknowledges hearing difficulties until the first contact with a hearing-health professional [9].

Developing easily accessible and affordable mobile health (mHealth) solutions in audiology would promote broader and faster access to diagnosis and health services, fostering an early rehabilitation and reducing the impact of HL on the individual. It is estimated that 55% of the global population and 87% of the European population will use internet services on a mobile device by 2030 [10]. It is thus evident that mHealth solutions have the potential to significantly impact behaviors in the population. Current studies indicate that the use of tablets and smartphones, including mHealth apps, is steadily increasing among adults aged 65 and older [11–13], who are even reported to be the fastest growing population of smartphone users [11]. In the last decades various mHealth applications have been developed for ear and hearing assessment [14–18] and for HA rehabilitation [19,20], in addition to tele-audiology services [21] and hybrid clinics [22]. The easiness of use and accessibility of mHealth solutions can have an impact on self-awareness and recognition of hearing loss and foster knowledge and use of in-person services [19,23,24]. Moreover, mHealth solutions show the potential to promote more equitable healthcare in low- and mid-income countries, where access to healthcare facilities and professionals is limited [16,25]. Finally, mHealth applications that are quick and easy to use in everyday life have the potential to provide clinicians with important information at both the diagnostic and intervention phases and can help exploring and understanding daily experiences of the user with HL and facilitate more timely responses [19].

A hearing mHealth app might become the first contact with a hearing diagnostic service and should therefore motivate individuals in need to seek professional help in a personalized manner, in order to maximize the impact of early health services on the population. However, even though professional support might be recommended given the hearing test result, many users might still be hesitant to seek help. This is particularly true among the older adult population, where HL can be slow and gradual and is often considered a natural aspect of ageing. Additionally, individuals may not be aware of the rehabilitation options and hearing healthcare services available to them, or how to access them [4]. Some users may even choose to ignore indication to seek help for various reasons, such as low awareness of HL or stigma associated with HA use. In addition, individuals high on neuroticism and low on agreeableness might generally distrusts external advice. Acquiring more information about users' personal characteristics and generating their individual profiles can inform the creation of targeted recommendations, particularly for those users who need more convincing incentives in order to take action. These recommendations can aid skeptical individuals in their decision-making process. For example, repeated feedbacks on daily hearing tests could increase self-awareness of HL and HA simulators could promote positive expectations toward HA. Moreover, information about users' personal characteristics collected by a hearing mHealth app could later assist clinicians in providing personalized counselling. It follows that the assessment of a person with hearing difficulties needs to go beyond the simple quantification of HL, and should also take into account individual characteristics that have been shown to influence the readiness of individuals to seek professional help [4,26].

Predictors of hearing help-seeking and hearing aid uptake

Hearing help-seeking can be seen as a first crucial step towards the decision to uptake a HA [26]. Help-seeking takes place after a contemplation stage [3], where a listener in need is initially ambivalent about making changes. Help-seekers would then prepare (seek information, plan) and take action [3] towards a change in behavior and attitude, namely consulting a healthcare professional about their hearing difficulties [4,27]. Acknowledgement and acceptance of hearing difficulties and their impact on everyday life have been discussed as the most important predictive features of hearing help-seeking, as well as later HA uptake [5,6,26,28,29]. In the older adult population HL is frequently perceived as part of the natural ageing process and other health issues are prioritized for treatment [4,30]. Even when HL is identified, individuals might reject the use of a HA due to expected costs, stigma and negative stereotypes [5,7,28]. However, a positive attitude towards HA [26,27], high expectations to benefit from them [26,29] and perceived self-efficacy in their daily management [17,18] were shown to promote help-seeking and HA uptake. Other relevant covariates that have been identified in the literature are personal attitudes, beliefs, and personality traits. Individuals who are more prone to seek help and successfully uptake a HA show higher internal locus of control [2,17] self-efficacy [31,32], and agreeableness, as well as lower neuroticism and openness [2]. Altogether such individual characteristics refer to a general self-confidence in the ability to cope with critical situations, good acceptance of others' suggestions and recommendations, as well as less susceptibility to shame and embarrassment.

Given the wide range of traits and behaviors that have been reported in the literature to be associated with hearing help-seeking and HA uptake, predictive models can be developed that take into account the multifaceted nature and association patterns of these traits and behaviors. Machine learning models are built upon a large number of predictors simultaneously, usually leading to more accurate predictions than univariate or smaller models that take into account only a limited number of predictors. These models can capture complex and non-linear relationships between the outcome and its predictors. In addition, when combined with cross-validation approaches, they draw more robust conclusions and generalize to new data. Currently, the use of machine learning to support healthcare applications is rapidly growing, particularly in the areas of predictive analytics, diagnosis and treatment, personalized medicine, clinical decision support, and population health management [33]. Machine learning algorithms and feature importance analysis can be used to identify the most relevant predictors of hearing help-seeking from the many features hitherto reported in literature. In an mHealth context, it is crucial to limit the number of features to a small set of key predictors to create a concise and efficient assessment battery.

Rationale and objective

The present study aims at identifying the most important predictive features for hearing help-seeking, aiming to design an individual profiling module for a hearing mHealth app. Such module will categorize help-seekers versus non-seekers and ultimately inform the design of targeted or even personalized treatment recommendations. For this purpose, data from a large number of questionnaires and tests covering different hearing-related and psychological characteristics, together with multiple assessments of a hearing screening test were collected in a longitudinal study that simulated a hearing mHealth app. To target potential users of future apps of this kind, the study was addressed to individuals with subjectively perceived hearing difficulties who had not yet been compensated with HA. We selected 25 assessment tools based on an extensive literature review of covariates of hearing help-seeking and HA uptake. From these, a comprehensive set of 83 features were derived. We used supervised machine learning algorithms to predict the readiness to seek professional help, as assessed at the end of the study and after two months as self-reports on intention to seek help. Feature importance analysis was used to narrow down the large number of features and identify a small set of key traits to predict hearing help-seeking behavior. Based on these results, we aim to derive suggestions for the implementation of a profiling module as a short and concise assessment battery, that can be performed after an audiological test. This could be included in existing or future mHealth apps in a modular manner. Knowledge of a user's propensity to seek help can be used to provide specific recommendations to encourage the use of hearing healthcare services. Ultimately, our aim is to provide clinicians and mHealth app developers with relevant knowledge about personal characteristics that are helpful in promoting hearing health by encouraging the uptake of hearing healthcare services and HA when needed.

The following research questions guided our study design and analyses:

RQ1. Which machine learning model can best predict help-seeking and categorize individuals into help-seekers versus non-seekers?

RQ2. Which hearing-related and psychological features are most relevant to classify individuals into help-seekers versus non-seekers?

RQ3. How can feature importance measures inform the design of targeted recommendations for users of a future hearing mHealth app?

Methods

Participants

Older adults above 50 years of age were recruited between August 2021 and August 2022 through the online platform Ebay's minijob announcements, the university intranet and via mailing list services of several German universities' Guest-Audience and Senior-University programs. Inclusion criteria were: subjective reports of hearing difficulties in daily life, ownership of and ability to use a smartphone, and good command of the German language. The exclusion criterion was the use of hearing aids. A total of 192 individuals were enrolled in the study. Seven participants dropped out during the study, resulting in a completion rate of 96%. The final dataset included N_1 = 185 participants, 106 females and 79 males (0 diverse), aged between 47 and 82 years, with M_{age} = 63.1 and SD_{age} = 6.5. One participant was below 50 years of age (47) but was nevertheless included in the final sample, given that this value only slightly deviated from the planned age threshold. Out of the N_1 = 185 participants who completed the study, N_2 = 131 attended the follow-up questionnaire. A descriptive summary of participants' socio-demographic characteristics is provided in Table 1.

Table 1. Main socio-demographic characteristics of the participants.[a]

Characteristic		***n* (% of total)**
Age group (years)		
	47-60	70 (37.8)
	61-70	83 (44.9)
	70-82	32 (17.3)
Sex		

	female	106 (57.3)
Duration of hearing difficulties (years)		
	0 - 1 2 - 5 6 - 10 10 – 21	54 (29.2) 89 (48.1) 23 (12.4) 19 (10.3)
Presence of Tinnitus		
		65 (35.1)
Previous doctor consultation for hearing difficulties		
		89 (48.1)
Presence of visual problems		
		134 (72.4)
Occupation status		
	employed	67 (36.2)
Monthly income (Euros)		
	< 1.500	62 (33.5)
	1.500 - 2.500	65 (35.1)
	2.500 – 4.000	42 (22.7)
	>4.000	16 (8.7)
Residential environment		
	countryside	2 (1.1)
	little town	29 (15.7)
	suburbs	48 (25.9)
	city	106 (57.3)
Self-estimated noise level at home		
	low	67 (36.2)
	moderate	113 (61.1)
	high	5 (2.7)

[a] These data were acquired during the baseline assessment through a self-report questionnaire.

Procedure

Study overview

Interested participants contacted the study administrator via email and received extensive written information about the purpose of the project, study design, length of participation and remuneration, possibility to withdraw participation at any time, data protection, management and storage. The study design, implementation, data collection, analysis and storage were conducted in accordance with current literature on ethics considerations in the context of mobile and mHealth apps [34,35]. Security and privacy recommendations were also adhered to. The study supported the autonomy of participants by providing extensive informed consent, which was given both as a separate written document prior to enrolment and within the formr online framework (see below for details). Debriefing was included, and participants were invited to provide feedback at the end of the study. No direct risks associated with the study design were identified, and privacy risks were accounted for through appropriate data management and data protection concepts for all software and platforms used. Personal information collected during the study was pseudo-anonymized with a written coding list stored in a closed locker accessible only to the study administrator. This coding list has been destroyed at the end of data collection, therefore the data is completely anonymized since. Data collection happened between September 2021 and September 2022. Participants were remunerated with 10 Euros per hour. A further incentive for study participation was that participants received a written feedback on their daily hearing test results. It has been clearly stated that a medical diagnostic is not provided in the study. Communications with the participants took place exclusively via email and SMS. A pilot study was conducted with a young (23 years of age), healthy female participant in August 2021 to evaluate the usability and technical functionality of the mobile study.

The study was conducted entirely online on personal mobile phones of the participants to approximate the experience of using an app. Only seven participants used their computers due to technical difficulties with their smartphones. Data collection was carried out using formr, an open-source web-based application programming interface for the R language that creates automated studies [36]. In formr different questionnaires and tests (see section Assessment) were linearly chained together as modules of a so-called run. In formr a run reproduces the desired design and can be assessed by users through a specific link. The software formr first provides a unique study link to the run, that was shared via email with enrolled participants. Upon accessing this link, participants were assigned a unique visitor session in formr and provided with a second individualized link, based on web cookies. The unique visitor session prevented users to provide multiple entries for the same survey. This unique session code enabled the anonymization of the data within formr and was for the time of the study stored in the aforementioned written coding list, where the participants' names and session codes were recorded. The Customer Communication Platform Twilio [37] was used to send the individualized study link to the participants through daily SMS reminders. Automated SMS delivery was initiated via the "external link" module in formr, which uses REST API (Representational State Transfer Application Programming Interface) to connect to Twilio. REST API allows a software program (in this case formr) to expose functionality and data to other programs (Twilio) in a consistent and secure format, ensuring privacy and data protection. With the individualized link received via SMS, participants could perform the study on their own smartphone's browser. For the seven participants who completed the study on their computer, the daily SMS were sent to their personal mobile phones as reminders. The individualized link was additionally sent via email at the beginning of the study, to allow these participants to access the study via their computer browser.

A detailed list of all assessment tools and their references, as well as the derived features for analysis, is provided in Table 2 and Table 3. Each assessment tool was implemented as a survey in formr. The majority of the surveys included in the study have been implemented following the paper-pencil version that was retrieved from the literature. The surveys that have been developed or adapted specifically for this study can be shared upon request. The items assessed can be inferred from the features provided in Table 2 and 3. Submission of each survey was possible only after all mandatory questions had been answered. After submission of one survey, the study advanced automatically to the following questionnaire or test planned in the formr run. Users were not allowed to go back and modify their answers after submission.

The total assessment time of eight hours was distributed across the working days of three consecutive weeks, with an overall daily active participation of approximately 30 minutes. The study design has been detailed in the Multimedia Appendix 1. Participants could select the starting date of the study, to ensure that the assessment could be easily integrated into their personal schedule. The first week (baseline assessment) included one measurement time-point per day (requiring approximately 20-30 minute to complete), which could be performed at any preferred time. On the first day, participants received an email with the study link and their unique participant code. After opening the study link on their browser, each participant has been given the possibility to read again a summary of the data protection conditions in the first page of the study. Secondly, they had been asked to provide telephone number and email address, which were then stored in formr and used for the automatic text reminders. They would then receive an automatic SMS with the individualized link to the study, through which they could begin the assessment. From day two to day five, participants received an SMS with the link to the study at 7 am, but they had been previously informed that they could perform the tasks at any preferred time during the day. An email reminder was sent at 7 pm in case participants didn’t access the study link by that time.

The second and third week included two measurement time-points per day of approximately 15 minutes each. The longitudinal assessments were prompted via SMS at 7 am and 7 pm. Participants were instructed to access the study at the earliest convenience after waking up and before going to bed, thus allowing them to accommodate the study to their daily schedules. In the morning, after clicking on the link they received via SMS at 7 am, each participant was first presented with some questions on baseline mood and sleep quality. They were then required to click on a second link embedded in the following survey page, that redirected them to the hearing assessment. Finally, participants were asked again to report on their mood after receiving a feedback on their hearing performance. In case the participant forgot to access the link and perform the study, an email reminder was sent at 1 pm. If, after the reminder, the participant still didn't take part in the study, the session was set as incomplete. The study administrator had to manually allow the participant to move to the next measurement time-point (in this case the evening assessment) within formr. This same assessment scheme was repeated in the evening. Here the SMS with the link to the study was sent at 7 pm and the email reminder at 11 pm.

At the end of the study, participants were asked to provide consent to be contacted after two months for a voluntary (and non-remunerated) follow-up online questionnaire. Those who provided their consent received an email with a link to a single formr survey that required less than 5 minutes to complete. The short survey consisted of two multiple-choice questions and was completed by 71% (N_2 = 131) of the participants. Individuals were asked to report again on their action to seek professional help following the feedbacks received during the study and were asked to indicate whether the study participation improved their awareness towards hearing difficulties.

Ethics approval and data availability

The study plan and data management have been approved by the Research Ethics Committee of the Carl von Ossietzky Universität Oldenburg. We, as single researchers and as a team, are committed to and supporting Open Science practices. We have therefore published a preprint of the manuscript on medrxiv.org and share data analysis scripts via Zenodo (please refer to the data availability statement for the links). The data will be shared with interested researchers upon request, since the dataset will be used for further projects and cannot be fully shared at the current time point.

Baseline assessment

During baseline assessment, cross-sectional data from a comprehensive set of 25 questionnaires and tests was collected. The questionnaires and tests were distributed on five consecutive days in order to maximize study compliance and avoid priming effects on different questionnaires. Below is a concise summary of the measured predictors (features). We refer to Table 2 and Table 3 for a complete list of the assessment tools used in the study. We selected questionnaires and tests that have been previously used in studies investigating their association with hearing help-seeking and HA uptake (as cited in the introduction and in the Tables below). If the tools included in the study were not previously used in similar literature, we explained our rationale for their selection in the following sections.

Assessment of hearing-related features

Firstly, the assessment included self-reports of participation and perceived handicap, focusing on self-reported hearing difficulties, consequences of HL and social life participation. Additionally, attitudes towards HA were evaluated with questionnaires on HA expectations and stigma. Hearing-related personality traits were also taken into account. Noise sensitivity was measured as a personality trait which was shown to be related to affect and neurosensory processing [38]. We further assessed hearing habits, aiming to gather more information about sound sensitivity and individuals' sound preference profiles [39]. Finally, hearing health literacy was assessed as well, since the ability to search, find and understand information related to hearing health was shown to be associated with better self-management of HL [31].

Assessment of psychological, general health and socio-demographic features

Personality traits (the Big Five [40]) were shown to be associated with help-seeking and HA use, and were therefore included in the baseline assessment. Anxiety and depression were also measured, given their frequently demonstrated associations with HL [4,6], together with loneliness, which is seen as consequence of untreated HL [3]. We further assessed optimism and sensory processing sensitivity, which refers to an individual's disposition to perceive and process stimuli (including auditory ones) more intensely than the average population [41]. Attitudes and beliefs like locus of control and self-efficacy were included as well for their association with help-seeking and HA use. The belief that HA are associated with old age and infirmity is often a barrier to HA uptake and use [4], therefore attitude towards ageing was assessed as well. Perceived stress was measured too, since high levels of stress that are related to daily life, work or social situations may boost help-seeking behaviors [3]. General health was assessed given its predictive role for different steps of the HA uptake path [4,6]. For completeness, we also measured cognitive abilities (crystallized and fluid intelligence), despite discordant findings on associations between cognition and HA uptake [4,6,16]. Lastly, participants were requested to complete a comprehensive questionnaire on socio-demographics.

Longitudinal assessment of hearing and affect

This micro-longitudinal assessment accounts for potential daily fluctuations of hearing performance and affect, which might depend on particular daily events and states. The affect questionnaire included 14 items in line with the Circumplex Model of Affect (Russel, 1980 [42]). Eight items were related to negative affect and six to positive affect [43]. The items are listed in the Multimedia Appendix 1. The affect questionnaire was presented before and after the hearing test in order to assess mood at baseline and after receiving feedback on the hearing test, respectively.

Hearing performance was assessed with the Digit Triplets Test (DTT) [44,45] of the Hörzentrum Oldenburg gGmbH. This widely used screening instrument [46] measures speech intelligibility in noise by means of the Speech Recognition Threshold (SRT), which indicates the Signal-to-Noise Ratio (SNR, difference between speech and noise level) at which the participant reaches 50% speech intelligibility. SRT measures obtained with the DTT show high correlations (higher than r = .70) with pure tone average (PTA) measures, while being relatively robust against changes in presentation level. Moreover, the DTT has shown to be robust to ambient noise levels outside of audiometric booth environments [47]. Together with its low linguistic and cognitive demands [47], the DTT appears to be suitable for mobile, remote self-test based screening of hearing abilities in the older population. Smartphone-based DTT has also shown the potential to provide widespread access to hearing screening in developing countries and across different socio-economic strata [48]. After completing the hearing test, participants received a feedback on their performance in the form of a traffic-light color, where green indicated good performance (SRT < -7.1 dB SNR), yellow indicated intermediate (-7.1 >= SRT < -5.1 dB SNR), and red reflected poor performance (SRT >= -5.1 dB SNR) [45,49]. The appendix provides detailed information on the hearing test (Multimedia Appendix 1) and its feedback (Multimedia Appendix 2). Participants performed the hearing test online on their personal smartphone and headphones. Three participants used loudspeakers as headphones were not available to them. Calibration of the hardware equipment was not possible due to the remote assessment. However, SRT estimation is relatively robust against changes in presentation level and no exact calibration is needed [50]. Moreover, the use of different types or quality of headphones have shown no impact on test reliability [46,48]. Each hearing test began with a signal adjustment trial meant to set the stimulus at approximately 65 dB SPL. The participant was presented with a digit triplet in noise and asked to “adjust the volume to hear both the digits and the noise clearly”.

Table 2. Assessment of hearing-related features (baseline and *longitudinal* assessment).

Domain	**Predictor (Assessment tool)**	**Feature for machine learning**
Participation and Handicap		
	Self-reported hearing difficulties (Speech, Spatial and Qualities of Hearing – SSQ, Kiessling, Grugel, Meister, & Meis, 2011)	- speech hearing scale - spatial hearing scale - qualities of hearing scale
	Consequences of hearing loss (Hearing Handicap Inventory for the Elderly and Adults - HHIE/A, Nuesse, Schlueter, Lemke, and Holube, 2020)	- social consequences of hearing loss scale - emotional consequences of hearing loss scale
	Social life participation (Social Network Index, Cohen, 2017, adapted from the Psychology department, University of Oldenburg)	- social network diversity score - number of people score - number of nets score
Attitude towards hearing aids		
	Hearing aid expectations (Expected Consequences of Hearing-Aid Ownership – ECHO, Kiessling et al., 2011)	- hearing aid expectations (global score)
	Hearing aid stigma (Attitudes towards Loss of Hearing Questionnaire - ALHQ v. 3.0, Saunders et al., 2005)	- denial of hearing loss scale - negative associations scale - negative coping strategies scale - manual dexterity and vision scale - hearing-related esteem scale
Hearing related personality traits		
	Noise sensitivity (Weinstein Noise Sensitivity Scale – WNSS, Zimmer and Ellermeier, 1997)	- noise sensitivity (global score)
	Hearing habits (Sound Preference & Hearing Habits Questionnaire – SP-HHQ, Meis et al., 2018)	- noise annoyance factor - sound quality factor - noise sensitivity factor - unpredictable sounds factor - openness factor - warm sounds factor - environmental sounds factor
	Hearing health literacy (European Health literacy questionnaire, short form - HLS-EU-Q16, Jordan & Hoebel, 2015; with additional 9 internally-developed items)	- health literacy (global score)
Hearing performance		

	Hearing performance (SRT) (Digit Triplet Test – DTT, Buschermöhle, Wagener, Berg, Meis, & Kollmeier, 2014)	- SRT mean - SRT standard deviation
	Hearing feedback type	- intermediate (% yellow feedback) - poor (% red feedback)
Others		
	Hearing-related socio-demographic data (Socio-demographic questionnaire, developed for this study from the authors)	- hearing difficulties: presence - hearing difficulties: duration (years) - hearing difficulties: previous consultation with a health professional - tinnitus: presence - tinnitus: duration (years) - tinnitus: previous consultation with a health professional - motivation to seek professional help before the study - source of motivation to seek professional help before the study - motivation to seek professional help after the study - source of motivation to seek professional help after the study - general attitude towards hearing aids

Table 3. Assessment of psychological, general health and socio-demographic features (baseline and longitudinal assessment).

Domain	**Predictor (Assessment tool)**	**Feature for machine learning**
Personality traits		
	Big 5 (NEO-ffi, Borkenau & Ostendorf, 1993)	- neuroticism scale - extraversion scale - openness scale - agreeableness scale - conscientiousness scale
	Trait anxiety (Geriatric Anxiety Inventory, Gottschling, Segal, Häusele, Spinath, & Stoll, 2016)	- anxiety (global score)
	Trait depression (Geriatric Depression Scale, Yesavage & Sheikh, 1986)	- depression (global score)
	Optimism/Pessimism (Die Skala Optimismus-Pessimismus – 2, Kemper, Beierlein, Kovaleva, & Rammstedt, 2012)	- optimism (global score)
	Loneliness (De Jong Gierveld Loneliness Scale - DJG scale, Tesch-Römer et al., 2013)	- loneliness (global score)
	Sensory processing sensitivity (High Sensitive Person Scale - HSPS-G, Konrad and Herzberg, 2019)	- ease of excitation scale - sensory threshold scale - aesthetic sensitivity scale - hearing scale
Attitudes and beliefs		
	Health Locus of Control (Kontrollüberzeugungen zu Krankheit und Gesundheit – KKG, Lohaus & Schmitt, 1989)	- internal locus of control scale - society control scale - external locus of control scale

	(Internale-Externale Kontrollüberzeugung-4 Skala, Kovaleva et al., 2012)	
	Attitude toward ageing (Attitude to ageing questionnaire – AAQ, Laidlaw, Power, & Schmidt, 2007)	- psychosocial scale - physical scale - psychological scale
	General self-efficacy (Generalized Self-Efficacy Scale – GSES, Schwarzer & Jerusalem, 2003)	- general self-efficacy (global score)
Mood		
	Affect (Daily questionnaire on affect, developed for this study from the authors)	- positive affect pre-test mean - positive affect pre-test standard deviation - negative affect pre-test mean - negative affect pre-test standard deviation - positive affect post-test mean - positive affect post-test standard deviation - negative affect post-test mean - negative affect post-test standard deviation
	Stress (Perceived Stress Scale – PSS, E. Schneider, Schönfelder, Domke-Wolf, & Wessa, 2020)	- perceived stress (global score)
Cognitive functions		
	Figural Reasoning (Berlin Fluid and Crystallized Intelligence Task – BEFKI, Schipolowski, Wilhelm, & Schroeders, 2020)	- figural reasoning (global score)
	Vocabulary (Wortschatztest, Schmidt & Metzler, 1992)	- vocabulary (global score)
	Digital literacy (Technikbereitschaft -Kurzskala, Neyer et al., 2012)	- technology commitment (global score)
Others		
	General health (Fragebogen zum Allgemein Gesundheitzustand SF-12, Drixler, Morfeld, Glaesmer, Brähler, & Wirtz, 2020)	- physical health score - mental health score
	General Socio-demographic data (Socio-demographic questionnaire, developed for this study from the authors)	- age - sex - presence of visual problems - education degree - occupation (retired or working) - weekly working hours - monthly income - relationship status - monthly income of the partner - residential environment - household size

Outcome measures

Classification of participants into help-seekers and non-seekers was based on self-reports of planned actions to seek professional help for their perceived hearing difficulties and their motivation to seek help. These variables, as retrieved at the end of the study and at follow-up, were chosen as outcome measures for the supervised machine learning (see below). This information was also assessed at the beginning of the study. Three different classifications will be considered as outcome measures: action to seek help (at study end), action and motivation to seek help (at study end) and action to seek help at follow-up. The distribution of participants along the outcome classes considered is summarized in Table 4.

Action to seek help

Help-seeking (preparation and action [3]) was assessed at study end with the question: "Given the feedbacks of this study regarding your hearing performance, have you made an appointment with one of the following doctors or a hearing care professional, or are you planning to do so?" (followed by a list of hearing professionals). This variable was used to create two outcome classes:

- *action* class (n_{11}= 64): participants who were planning to seek professional help in the near future or had already taken an appointment;
- *no action* class (n_{12} = 121): participants not ready to take action, who did not plan to consult a hearing health professional in the future.

Action and motivation to seek help

A second outcome measure was taken into account to further differentiate the *no action* class, in order to provide further insight for the design of targeted recommendations in a hearing mHealth app. Information on readiness to take action was combined with the reported motivation to seek help at the end of the study. Motivation was assessed through the question: "How motivated are you at the moment to seek help regarding your hearing problems?" (1 = not motivated at all, 7 = very strongly motivated). The answer spectrum was binarized by means of median split to create the following outcome classes:

- *no action & high motivation* class (n_{13} = 47): participants not ready to take action with high motivation, who might particularly benefit from personalized and tailored recommendations;
- *no action & low motivation* class (n_{14} = 74): participants not ready to take action who reported low motivation to seek help;
- *action* class (n_{15} = 64): participants ready to take action, regardless of their motivation level. This class was not further divided with respect to motivation, since this would not result in different recommendations. Moreover, data exploration revealed that only seven individuals out of 64 in this category reported low motivation.

Action to seek help at a follow-up

The voluntary follow-up questionnaire (completed by N_2 = 131 participants) included a question on the intention/action to seek help following the feedbacks received during the study. The answers (given as four multiple-choice answers) were binarized to achieve a class allocation comparable to the first outcome measure:

- *action at follow-up* class (n_{21} = 52): participants who reported having completed an appointment with a hearing professional, who had an appointment scheduled but not completed or who were planning to seek help in the near future;
- *no action at follow-up* class (n_{22} = 79): participants who did not plan to consult a hearing health professional;

Table 4. Absolute class-wise frequencies of observations at 2-month follow-up across the three outcomes considered at the end-of-study.[a]

	action at follow-up	**no action at follow-up**	**no follow-up data**	total

action				
	33	12	19	64
no action & high motivation				
	12	22	13	47
no action & low motivation				
	7	45	22	74
total				
	52	79	54	

[a] The three outcomes considered are: action to seek help (N_1=185); action and motivation to seek help (N_1=185); action to seek help at follow-up (N_2=131). Additionally, the table provides an overview of those participants who did not complete the follow-up questionnaire.

Statistical analysis

Data pre-processing

Data analysis was performed with the R Software for Statistical Computing [51]. Raw data from all questionnaires was imported from the online platform formr to R environment using the dedicated package formr [52]. For each questionnaire or test presented at baseline, global scores were computed and considered as features. If both global scores and scale scores were available for a given assessment tool, only the scale scores were maintained if considered differentially relevant for the outcome. Hearing test results were received from the Hörzentrum Oldenburg and imported in R as .eml files. The performance feedback category (green, yellow, red) was additionally extracted and stored for each raw SRT result. Due to the particular implementation of the study in formr, participants could perform the hearing-test more than once at each measurement time-point. Whenever this happened, only the last SRT result at a given time-point was kept for analysis. This led to a removal of 3.9% of the raw SRT results. The longitudinal data on daily mood and hearing performance was summarized into individual means and standard deviations. The summarized longitudinal data was merged with the cross-sectional data, resulting in a wide-format data frame including 83 features. Table 2 and 3 provide a complete list of features considered for analysis.

Completeness rate for the baseline questionnaire was 100%, while missing data occurred for the longitudinal measures of hearing performance and affect. A complete set of 20 SRT results was collected for 43.8% of participants, while at least 15 SRT results were obtained in 95.1% of cases. Missing hearing data at a specific time-point was considered Not Available (NA). Where an SRT result was missing, the respective feedback and measures of affect at post-test were set to missing as well (NA). By visual inspection of the individual SRT distributions, some specific outlier patterns were identified. 14 participants showed a much larger SRT result at the first measurement, which qualified as outlier following the interquartile range rule. These large SRT values (indicating poor performance) were considered to be caused by misunderstanding of the hearing test instructions and were therefore set to NA. The respective feedback category and measures of affect where however not set to NA. This is because despite the unreliable SRT value, participant’s mood could still have been affected by the feedback received. With respect to daily affect measures, two participants provided no data at measures of post-test affect, such that summary measures could not be computed. In these cases, mean imputation technique was applied: the sample mean and sample variance for negative and positive affect at post-test were imputed to replace missing values.

Machine learning

The data was fed into three machine learning algorithms for supervised classification. We chose

Naïve Bayes (NB), Random Forests (RF) and Support Vector Machines (SVM) among other classifiers to cover a wide range of model complexity (from simple models like NB, to more complex and non-linear ones like SVM). The algorithms have been implemented on R using the mlr package [53], following the approaches described in [54] and [55]. Given the presence of three different outcome measures, the same analysis steps were carried out in parallel for each outcome, with a slight difference in the input features included in the analysis. For the first outcome (*action to seek help*), data on motivation at pre-study was kept in the feature space, while motivation to seek help at the end of the study was removed. Same applied to the third outcome (*action to seek help at follow-up*). Differently, for the second outcome (*action and motivation to seek help*), all data on motivation at pre- and post-study was removed from the feature space. Due to the relatively small dataset and imbalanced classes we chose not to split the data into training- and test-set but to use Cross-Validation instead. CV divides the training set into *k* equally sized parts and considers the *k*th part as test set and the *k*-1 parts as training set at each iteration. Model results are then averaged across all iterations. Implementation details of the three algorithms are summarized next.

Naïve Bayes classifier (NB)

This algorithm uses Bayes' rule to predict the probability of an observation to belong to one of the outcome classes given its discriminant function values. Given the prior probability, the likelihood and the evidence for each observation, the relative posterior probability for each class is computed. The single observation will then be assigned to the class with the highest relative posterior probability [54]. Two strong assumptions made by NB algorithms are: the normal distribution of continuous features (or predictors) and the independency of these features. Model performance will suffer in case of violation of these assumptions [54]. In the present implementation, after training the algorithm, repeated 10-fold CV was used to evaluate the model's performance. A step-wise approach was used to select the appropriate number of CV repetitions necessary to achieve accurate and stable performance estimates (50, 100, 150 and 300 CV repetitions).

Random Forest (RF)

Tree-based methods use recursive binary splitting to stratify the features' space in smaller, non-overlapping regions used for classification. At each iteration of the tree-building process, the algorithm selects among all features the one that best splits the data into two branches according to a specific question or rule (node) [54]. The process iterates until a stopping criterion is met and final regions (leaves) are identified. In a classification problem, the mode of the training data within a region is used for prediction: each observation is classified to the majority class within the leaf to which it belongs [56]. Trees are easy to interpret but they lack of predictive power, since they tend to overfit the training data. Approaches like RF can be used to improve prediction accuracy. RF is a non-linear method that involves the generation of multiple uncorrelated trees from different bootstrapped training sets, obtained by sampling with replacement from the original data. The final predicted outcome is retrieved from aggregating the prediction of all built trees and selecting the most frequent or modal prediction [54,56]. This algorithm requires to tune a set of hyperparameters which control the learning process and are selected (or tuned) by the algorithm to obtain best performance. The following hyperparameters were considered:

- Ntree: number of trees to include. This value is usually fixed at a computationally reasonable value rather than tuned [54]. Ntree was set to 800.
- Mtry: number of features to randomly sample at each time. A popular value is given by $\sqrt{p}$ (where p = the number of predictors) [57]. Different search spaces were explored, with Mtry ranging between 1 and 15
- Nodesize: minimum number of cases to be included in a leaf. Different search spaces were explored, with Nodesize ranging between 1 and 20.

Tuning the algorithm and finding the best hyperparameter combination requires to define an optimization algorithm, or search strategy, and evaluation method. We used grid-search with 10-fold CV resampling. To evaluate model performance, nested CV was applied. In this approach, an inner loop tunes the hyperparameters and an outer loop evaluates a wrapped learner, which comprises the classification task, the learner type (RF) and the hyperparameter tuning process. Here a 5-fold CV was applied as outer resample strategy.

Support Vector Machine (SVM)

The SVM algorithm iteratively identifies a hyperplane that separates labelled classes also in case of non-linear data distributions. It does so by adding an extra dimension to the data, which is found through the kernel function, a mathematical transformation of the data. The hyperplane is defined as a surface that has one dimension less than the number of variables in the dataset. The position of the hyperplane depends on the position of the support vectors, which are training-set cases that define the class boundaries [54]. The optimal hyperplane is found by maximizing its margin, which is the region around the hyperplane that touches the fewest training observations. In fact, the distance of a training case from the margin can be viewed as a measure of correctness of its classification [56]. In case the algorithm needs to separate more than two classes, several models are built and compared to find the one that best predicts new data. SVMs are computationally expensive, but tend to perform very well on a variety of tasks conducted on non-linearly separable classes. Additionally, the algorithm has the advantage to make no assumptions on the features' distributional properties [54,56]. Similar to RF, SVM requires hyperparameter tuning. The following hyperparameters were considered:

- Kernel: type of kernel function used to identify the hyperplane. Polynomial, radial and sigmoid functions were included in the search space [54].
- Degree: shape of the decision boundary (in case of a polynomial kernel). The search space was limited to values from 1 to 5 to avoid the risk of overfitting [54].
- Cost: penalty for having cases fall inside the margin. It is recommended to tune both cost and gamma (see next point) on the logarithmic scale [58], and a popular search space for cost is from 2^{-5} to 2^{15} [59].
- Gamma: influence of each case on the hyperplane. This hyperparameter search space was set to the popular range 2^{-15} to 2^{3} [59].

Nested CV was used as previously described for RF. An inner resampling loop (with 10-fold CV) was applied for hyperparameter tuning and an outer loop (with 5-fold CV) for performance evaluation.

Classification performance metrics

The algorithms were evaluated in terms of prediction accuracy on the test set which indicated the overall proportion of cases correctly classified by the model as compared to the observed outcome. However, class imbalance in the sample can negatively impact prediction accuracy, reducing its informativeness as performance measure [60,61]. Matthews Correlation Coefficient (MCC) (yardstick package [62]) was additionally taken into account to evaluate model performance. This coefficient improves over accuracy measures in case of imbalanced datasets [61,63] and can take values from -1 to 1, with 1 indicating perfect prediction, 0 chance prediction and -1 inverse prediction. As an additional metric we computed the confusion matrix (`calculateConfusionMatrix()`, mlr package). Its output provides the absolute number and the proportion of correct model predictions and misclassifications for each outcome class. For binary outcomes, the confusion matrix allows to estimate model sensitivity (i.e. accurately identifying seekers) and specificity (i.e. accurately identifying non-seekers). In the present study, obtaining high specificity is of particular interest in the context of a hearing mHealth app. Indeed, individuals with HL who are not prone to seek help are the main target population for tailored treatment recommendations and counselling.

Feature importance

After identifying the best performing machine learning algorithm, feature importance was considered. Each feature receives a coefficient of importance that indicates its contribution to model performance, regardless of the type of relationship (direction and linearity) between the feature and the outcome [64]. In RF, feature importance is model-dependent and it indicates how much the feature contributes in reducing node impurity. Importance values were retrieved by the function `getFeatureImportance()` (mlr package) applied on the RF model trained with the tuned hyperparameters. These importance results have the advantage of being inherent to the model and closely tied to its performance [65]. Conversely, there are no model-specific importance metrics available for NB and SVM. For these algorithms, the importance value assigned to each feature corresponds to the area under the ROC (Receiver Operating Characteristic) curve, which is computed from sensitivity and specificity measures [65]. The function `varImp()` from the caret package [66] was applied on the model trained with the function `train()` (caret package), after ensuring comparable performance with the same model previously trained in the mlr package.

Features with higher importance ranking will be considered for inclusion in the profiling algorithm, as they represent the most relevant predictors for telling apart help-seekers and non-help-seekers. No statistical criterion exists to determine which importance value threshold should be used to retrieve relevant features. Hence, three threshold values were inspected (the first 10, 15 and 20 features in their importance ranking order) and evaluated in terms of predictive accuracy and interpretability. Classification accuracy of these three feature sets were assessed on the outcome data obtained at follow-up. For this analysis, the dataset was reduced to N_2 = 131 participants who completed the follow-up questionnaire. The important features were fed into the best performing machine learning algorithm from the previous step.

Results

Machine learning classification performance

Predicting the action to seek help

A summary of the model-specific classification accuracy for the first outcome (*action to seek help*) is provided in Table 5. The three algorithms show similar overall performance accuracy estimates on the test set, correctly classifying about 67% to 70% of the cases in the full dataset (N_1 = 185). RF was the best performing algorithm with an accuracy of 70% and an MCC = .28, indicating that the model's prediction improves to about 20% over chance. By inspecting the confusion matrix (see Figure 2), we observed that RF shows high specificity, correctly classifying 91% of the cases belonging to the *no-action* class. The NB classifier (10-fold CV repeated 50 times) showed the best sensitivity compared with the competing algorithms, with 51% cases in the *action* class being correctly classified. RF and NB were selected for feature importance analyses, given the good predictive performance and high specificity of the RF, as well as highest sensitivity of the NB.

Table 5. Model-specific overall performance and class-specific classification accuracy rates for the first outcome measured at study-end.[a]

Action to seek help			
Model	**Hyperparameters**	**Overall performance**	**Class-specific**

				measures		classification accuracy	
	parameter	space	tuned	test accuracy	MCC	action (n_{11}=64)	no action (n_{12}=121)
NB							
				.66	.26	.51	.75
RF							
	ntree	800	800	.70	.28	.31	.91
	mtry	[5,15]	13				
	nodesize	[1,5]	1				
SVM							
	kernel		radial	.67	.23	.39	.82
	degree	[1,5]	4				
	cost	[2^{-5}, 2^{15}]	.01				
	gamma	[2^{-15}, 2^{3}]	69.1				

[a] Results are based on the full data set (N_1=185). For RF and SVM the table additionally shows the hyperparameter search space used in the resampling procedure and the tuned values used for model training and feature importance analysis. The NB model selected was 10-fold CV repeated 50 times. (NB: Naïve Bayes classifier; RF: Random Forest; SVM: Support Vector Machine; MCC: Matthews Correlation Coefficient)

Predicting the action and motivation to seek help

Table 6 summarizes the classification performance with respect to the second outcome (*action and motivation to seek help*), which includes three classes (see above). RF provided the highest accuracy with 55% and MCC=.30, as compared to NB (10-fold CV repeated 100 times) and SVM. However, the confusion matrix revealed that none of the three models was able to adequately tell apart individuals within the *no action* class who differ with respect to high versus low motivation. All algorithms can only correctly classify 2 to 25% of cases in the *no action & high motivation* class. Potentially, an improvement in classification accuracy could be achieved with a larger dataset in which the classes are better balanced, and with a more reliable and elaborated measure of the participant's motivation to seek help. In view of these limitations, the second outcome will not be considered for feature importance analysis.

Table 6. Model-specific overall performance and class-specific classification accuracy rates for the second outcome measured at study-end.[a]

Action and motivation to seek help								
Model	**Hyperparameters**			**Overall performance measures**		**Class-specific classification accuracy**		
	parameter	space	tuned	test accuracy	MCC	action (n_{15}=64)	no action & low motivation (n_{14}=74)	no action & high motivation (n_{13}=47)
NB								
				.50	.22	.47	.67	.25
RF								
	ntree	800		.55	.30	.55	.80	.09
	mtry	[8,10]						
	nodesize	[3,15]	800					
SVM								
	kernel		sigmoid	.49	.20	.53	.76	.02
	degree	[1,5]	1					
	cost	[2^{-5}, 2^{15}]	323					

	gamma	$[2^{-15}, 2^{3}]$	3.05×10^{5}					

[a] Results are based on the full data set (N_1=185). For RF and SVM the table additionally shows the hyperparameter search space used in the resampling procedure and the tuned values used for model training and feature importance analysis. The NB model selected was 10-fold CV repeated 100 times. (NB: Naïve Bayes classifier; RF: Random Forest; SVM: Support Vector Machine; MCC: Matthews Correlation Coefficient)

Feature importance

Predicting the action and motivation to seek help at follow-up

Feature importance was analyzed based on the RF and NB algorithms predicting the first outcome, *action to seek help* at study-end on the full data set (N_1 = 185). Each importance value signifies the feature's contribution to the model's performance. However, as detailed in the methodology section, RF and NB models calculate these coefficients differently. Consequently, ranking values were employed. In both models, the 83 features were initially ranked in descending order based on their importance values. Consequently, features with higher importance rankings are mostly relevant in predicting help-seeking behavior. Three sets of features among the most important ones were taken into account for subsequent analysis:

- the top-10 features indicated by the two models, resulted in a total of 12 best features;
- the top-15 features indicated by the two models, resulted in a total of 19 best features;
- the top-20 features indicated by the two models, resulted in a total of 28 best features.

Figure 1 shows all 28 features with their importance ranking values originating from the RF and NB models. The specific importance values for each feature are provided in the Multimedia Appendix 3. Next, the three sets of features (top-10, top-15 and top-20 features) were evaluated for their predictive performance and classification accuracy on the reduced dataset of N_2 = 131 participants who completed the follow-up questionnaire. NB and RF were trained on the three different feature sets for predicting the action to seek help at follow-up. Results are summarized in Table 7. They show that all feature sets provide good predictive performance and that the NB algorithm outperforms RF, with an overall accuracy ranging between 73% to 75% and an MCC between .43 and .47. Class-specific classification accuracy is comparable between NB and RF, with the action class correctly classified in 52 to 63% of the cases and the no-action class in 82 to 86% of cases. As can be seen from the confusion matrix in Figure 2, sensitivity and specificity measures are comparable for both algorithms predicting action to seek help at follow-up using the set of top-10 most important features. Sensitivity ranges from 55% (NB) to 58% (RF) and specificity from 84% (RF) to 65% (NB).

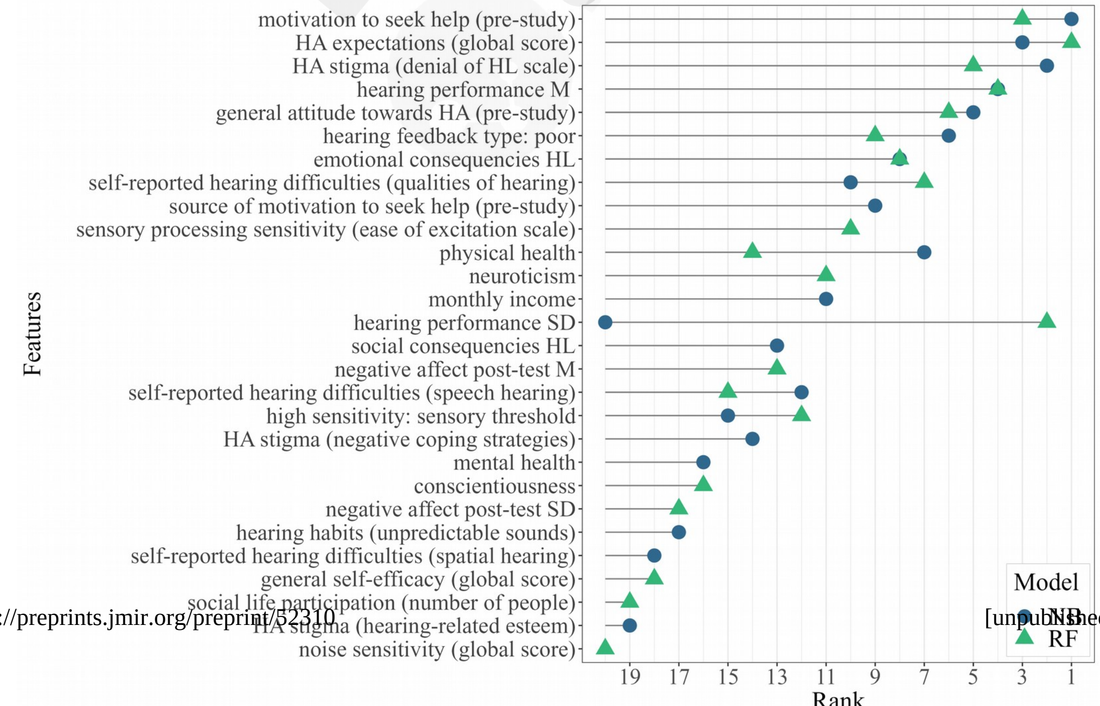


Figure 1. Ranking of the most important features to predict *action to seek help* at study-end. Importance ranking are shown for the 20 most important features for the two models (NB and RF) trained on the full dataset (N_1=185), as summarized in Table 5. A total of 28 features are arranged on the *y*-axis with respect to their average ranking between the two models. (HL: hearing loss; HA: hearing aids; *M*: mean; *SD*: standard deviation; NB: Naïve Bayes; RF: Random Forest)

Table 7. Model-specific overall performance and class-specific classification accuracy rates for the outcome measured at follow-up.[a]

Action to seek help at follow-up							
	Model	**Hyperparameters**		**Overall performance measures**		**Class-specific classification accuracy**	
		parameter	space	test accuracy	MCC	action (n_{21}=52)	no action (n_{22}=79)
Top-10 features (total 12)	**NB**						
				.74	.44	.55	.86
	RF						
		ntree	800	.73	.43	.58	.84
		mtry	[1,4]				
		nodesize	[1,5]				
Top-15 features (total 19)	**NB**						
				.73	.43	.58	.83
	RF						
		ntree	800	.73	.41	.56	.84
		mtry	[1,10]				
		nodesize	[1,5]				
Top-20 features (total 28)	**NB**						
				.75	.47	.63	.83
	RF						
		ntree	800	.70	.36	.52	.82
		mtry	[1,5]				
		nodesize	[1,5]				

[a] Results are based on the reduced follow-up dataset (N_2=131). The table shows the results for three sets of features (10, 15 and 20 most important features for the two models). For NB 10-fold CV repeated 50 times was used. For RF the table additionally shows the hyperparameter search space used in the resampling procedure. In contrast to the previous results (Tables 5 and 6), the tuned hyperparameters were not retrieved, as no importance analysis followed. (NB: Naïve Bayes

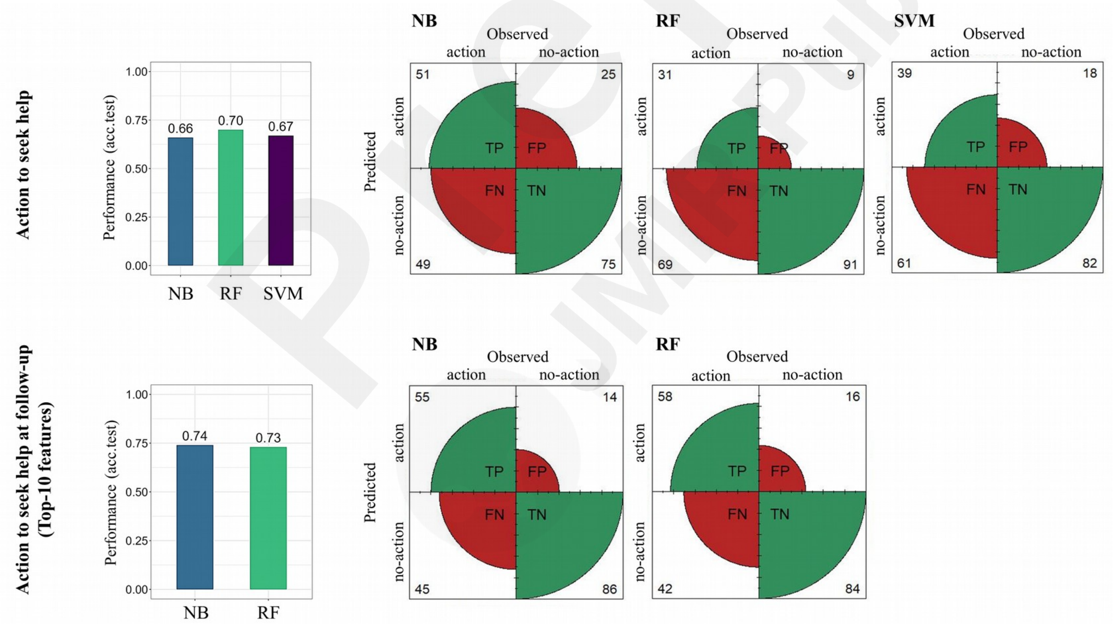
classifier; RF: Random Forest; SVM: Support Vector Machine; MCC: Matthews Correlation Coefficient)

Figure 2. Visualization of the model-specific performances (left) and relative confusion matrices (right) for the outcomes action to seek help (first row) measured at study end (N_1 = 185) and action to seek help at follow-up (second row) considering the set of top-10 most important features (N_2 = 131). The bar plots on the left show the model-specific performance for each outcome measured as prediction accuracy on the test set. Each confusion matrix plot illustrates the class-specific relative frequencies for each observed and predicted class, and provides a more immediate overview of each model's sensitivity (TP) and specificity (TN). (TP: True positive; TN: True Negative; FP: False Positive; FN: False Negative; NB: Naïve Bayes classifier; RF: Random Forest; SVM: Support Vector Machine)

Important features

The following is a brief description of 12 features ranked among the most important (previously named as the top-10 features for the two models) in the prediction of *action to seek help at follow-up*:

- *Motivation to seek help* and *source* of this motivation at the beginning of the study;
- Individual's attitude and expectations towards HA, including: *general attitude towards HA*;



expectations towards HA, as measured by the global score of the Expected Consequences of Hearing-Aid Ownership (ECHO) questionnaire [67], which assesses positive and negative expectations towards HA, expected services and costs, and assumptions about change in the personal image in case of HA use; and *stigma towards HA*, as measured by the Denial of Hearing Loss scale of the Attitudes towards Loss of Hearing Questionnaire (ALHQ–3.0) [68], which assesses acceptance of hearing aids and acknowledgement of HL;

- *Hearing performance* (mean SRT and its variability as measured by the DTT [44,45]) and *percentage of negative feedback received* (indicating poor performance);

- Perceived consequences of HL, including: *emotional consequences of HL*, as measured by the corresponding subscale of the Hearing Handicap Inventory (HHIE/A) questionnaire [69]; and *self-reported hearing difficulties*, as measured by the Qualities of Hearing subscale of the Speech, Spatial and Qualities of Hearing (SSQ) questionnaire [67], which addresses recognition, perceived clarity and naturalness of everyday sounds, as well as listening effort experienced in different hearing contexts;
- *High sensory-sensitivity personality*, as assessed through the Ease of Excitation subscale of the High Sensitive Person Scale (HSPS) questionnaire [70], which assesses emotional reactivity to physiological stimuli;
- Reported *physical health*, measured by the corresponding items of the Short-Form Health Survey-12 (SF-12) questionnaire [71].

As outlined in the methods section, ranking values up to 10 were arbitrarily chosen to identify the most important features. The following two features had slightly lower importance values (and received ranking value 11) but might provide further insights toward targeted counselling in a hearing mHealth app.

- *Neuroticism*, which refers to a predisposition of experiencing negative emotions [2] and was assessed through the corresponding items of the NEO Five-Factor Inventory (NEO-FFI) questionnaire [72];
- *Monthly income*, which was categorized with three cut-off values (less than 1.500 Euros, between 1.500 and 2.500 Euros, between 2.500 and 4.000 Euros and above 4.000 Euros).

Discussion

Key findings

The present study contributes to the identification of individuals' hearing-related, psychological and general health-related traits that predict the readiness to seek professional help for hearing loss (HL). Cross-sectional and longitudinal data have been collected in a comprehensive mobile study. Potential users of a future hearing mHealth app, namely individuals with subjective hearing difficulties, were classified into help-seekers and non-seekers by means of supervised machine learning algorithms. The trait measures used in this study were collected from previous literature investigating health care seeking, particularly in the audiological domain. From these, we derived a comprehensive set of 83 features to be used for prediction and profiling. The three algorithms taken into account (naïve Bayes, random forest and support vector machine) accurately predicted help-seeking behavior at the end of the study in 66 to 70% of cases. Especially, the random forest algorithm achieved high specificity, meaning that it was most successful in identifying individuals who might not intend to seek professional help. By selecting a subset of important traits revealed by our empirical feature importance analyses to predict hearing help seeking, the prediction accuracy for action to seek help at a two-month follow-up reached 75%. The present study identified the following features to be most important in the prediction of help-seeking behavior: perceived consequences of hearing loss in daily life, motivation to seek help, attitudes towards hearing aids, sensory-sensitivity, neuroticism, physical health and income. We conclude that these individual characteristics should be assessed in a profiling module that could complement the main auditory assessment module for hearing screening of existing or future mHealth apps. The degree of hearing loss, but importantly also its day-to-day variability were among the most important predictors, suggesting the need to perform repeated hearing assessments, which could be prompted by the app in different times of the day. To streamline the implementation of the profiling module in a mobile app, the questionnaires and subscales used to measure these important features should undergo item selection analysis to derive simple and short scales, yet reliable and valid. By incorporating a selected machine learning algorithm (random forest), the app can profile users into help-seekers or non-seekers based on the data collected by this short questionnaire battery. This information would complement the audiological data gathered by existing hearing screening/diagnostic tests, providing an informative user profile. The derived profile would guide the app in selecting the appropriate set of recommendations, optimizing an intervention on help-seeking behavior where needed. The results of the present study will also provide suggestions for the design of such targeted treatment recommendations. Ultimately, our aim is to provide clinicians and mHealth app developers with relevant knowledge to promote hearing health by encouraging the uptake of hearing healthcare services and HA when needed.

Which machine learning model can best predict help-seeking and categorize individuals into help-seekers versus non-seekers?

Three machine learning classifiers correctly predicted *action to seek help* at study-end in 66 to 70% of cases, clearly improving over chance prediction. This is a promising result considering the complexity of the prediction outcome. As discussed earlier, several individual factors can influence the decision to pursue hearing health care services and there can be discrepancy between contemplating, planning and taking concrete action [3]. Random forests showed best prediction accuracy and high specificity, while naïve Bayes showed the highest sensitivity. When predicting *action to seek help at follow-up* with the selected important features, the performance of random forest and naïve Bayes models improved up to 75%, despite the smaller dataset (N_2 = 131). Naïve Bayes showed higher predictive performance for this outcome. Overall, all models exhibited high specificity (ranging from 75 to 86%) and comparatively low sensitivity (31 to 58%). This could be attributed to the fact that the *action* class encompasses individuals who have sought professional help as well as those who are only considering taking action. Random forest showed high accuracy in identifying *no action* both at study-end and at follow-up and can be therefore considered the best performing algorithm in the present framework. Accurate identification of non-seekers is the most relevant performance outcome in an mHealth app to design targeted recommendations. Indeed, the envisioned profiling algorithm should be a system with high specificity that motivates and promotes help-seeking, especially in those cases where users would not spontaneously take action.

Which hearing-related and psychological features are most relevant to classify individuals into help-seekers versus non-seekers?

Hearing performance appears to be one of the most important features to predict help-seeking. The association between degree of hearing loss and help-seeking, as well as with hearing aid (HA) uptake, is well established in literature [4,9,26,28,29]. The present results also highlight – to our knowledge for the first time in the literature – the predictive role of intra-individual fluctuations of hearing performance, emphasizing the need to move beyond the traditional view of hearing as a stable neurosensory process [73]. The implementation of repeated daily measurements of hearing performance provides further insight on the impact of HL on the individual’s everyday life. In line with this, feature importance findings emphasize the relevance of self-reports on the consequences of HL. The assessment should consider self-reported listening effort in different contexts, as well as perceived handicap and emotional consequences of HL. Indeed, individuals who report greater negative impact of HL in their life are more prone to seek help and later uptake HA [9,27]. The individuals’ self-awareness of HL can be validated or improved by providing repeated feedback on hearing performance in a hearing mHealth app. As occurred at the follow-up survey, 85% of participants (out of N_2 = 131) reported increased awareness for their hearing abilities after receiving repeated feedback during the study. Finally, according to present results and previous findings [26,27,29], investigating stigma, attitude and expectations towards HA informs on individuals' readiness to seek help, as well as later uptake a HA. Stigma and negative stereotypes related to HA may deter individuals from seeking help and can represent a barrier to HA use [7,28].

Audiological factors emerged as the most important features. Nevertheless, other general health and psychological factors were also relevant in the prediction of help-seeking. In the present study, physical health was an important predictor for help-seeking, although the evidence on this relationship is discordant [26]. While people with better self-reported health were more likely to seek help [4,27], HA uptake was predicted by poor self-reported health [6]. Other important factors were related to the personality traits of sensory sensitivity and neuroticism. Individuals characterized by high sensitivity to sensory stimuli [41] and emotionally instable personality traits seem to perceive increased psychological discomfort following HL, even in presence of effective HA treatment [2]. Finally, income emerged as another important predictive feature. This is in line with evidence suggesting that higher socioeconomic status [9], higher income or pension earnings [3,27] and access to financial support [26] promote HA uptake.

How can feature importance measures inform the design of targeted recommendations for users of a future hearing mHealth app?

By assessing and analyzing the aforementioned hearing-related and psychological traits, the algorithm established here aims to profile the user as help-seeker or non-seeker. Completing this profile with complementary audiological test results, an informative picture of the user can be derived. Using this profile, clinical experts and intervention app designers could propose different sets of recommendations to assist individuals in their decision-making process in a targeted manner. We propose to first differentiate between profiled help-seekers and non-seekers, where the former should receive simple and straightforward recommendations only depending on their hearing status. For non-seekers, there is a need to design more specific and targeted recommendations based on information about the relevant characteristics to predict hearing aid seeking. Users who were profiled as determined help seekers could receive clear and concise guidance on the hearing care they needed. Those among them who should uptake a hearing device (given their audiological outcome), could benefit from additional information on available hearing-care services and professionals, in order to facilitate faster HA adoption rates. This would facilitate individual's perceived competence and autonomy, which are important predictors of hearing health-seeking behavior [74]. On the other hand, users with HL who are profiled as non-seekers should receive more elaborate, targeted recommendations to motivate and promote access to hearing care services. Recommendations for non-seekers should be further differentiated and designed depending on their perceived consequences of hearing loss, attitudes towards hearing aids, sensory-sensitivity, neuroticism, and potentially income. These recommendations could act as an intervention on modifiable predictive features like self-recognition of HL and attitude towards HA. For example, users profiled as non-seekers with good awareness and self-recognition of HL but negative expectations towards HA and low income, should receive a different set of recommendations than non-seekers with low self-awareness and high neuroticism.

Where HL is detected, the mHealth app could prompt repeated testing in different days and times of the day, and provide individual feedback on the performance compared to normative data. More detailed feedback on daily hearing performance could improve awareness on the hearing deficit. The app could also inform the user on the risks of an untreated HL and the benefits of early intervention through HA. Indeed, it has been shown that individuals are more likely to positively change their behavior when provided with actionable and meaningful information on their health status [75]. Where there is a need to promote positive attitudes towards HA, information could be provided on the wide range of devices available, as well as examples of successful peer cases. Knowledge on accessible financial support for HA by insurance companies could additionally promote HA uptake, given the predictive role of income. Furthermore, an implemented HA simulator in a hearing mHealth app could offer possibilities to experience improved listening conditions and promote positive expectations towards a HA. Elaborated information on HA technologies, like the benefits of noise control and noise reduction algorithms, could promote help-seeking by fostering knowledge of individuals who are more sensitive to environmental noise (high sensory sensitivity trait). The effectiveness of such recommendations could be further increased through targeting or tailoring communication. Targeted messages are designed for a specific population, while tailored communication is individualized to the person and was shown to be the most effective in promoting health behavior change [76]. Indeed, messages that are congruent to the personality traits of the audience are more positively evaluated, persuasive and generate more interest [77]. The predictive role of neuroticism for help-seeking behavior can be considered for efficient communication, both in the context of a hearing mHealth app and in clinical counselling. Individuals with high neuroticism trait are more susceptible to perceived disease [78] and are drawn to action by motives of safety and security [77]. For example, recommendations that target profiled non-seekers with low expectations towards HA can be differentiated depending on personality traits. To promote positive HA expectations in individuals with high sensory sensitivity traits, recommendations could focus on the benefits of HA related to noise control and noise suppression. However, such recommendations may not be effective for people who do not have this high sensitivity trait but who, for example, score high on neuroticism. Instead, they might be more convinced by recommendations that emphasize the risks of an untreated HL and the benefits of an early intervention through HA.

Limitations and future directions

The predictive performance of the machine learning classifiers could be improved in future studies with a larger dataset and more balanced classes. Classification accuracy could be further improved by including additional objective measures to complement participant self-reports. Continuous psychophysiological measurements (e.g. heart rate variability) could be included as further predictive features. This information could complement the longitudinal assessment of affect and better characterize potential changes in arousal before and after the completion of the auditory measurements. Note that multicollinearity as a potential statistical limitation was ruled out (the correlation plots are available in the Multimedia Appendix 4). Future studies might also benefit from longer follow-up to properly capture those individuals who took more time to take action to seek help. In the present study, measuring help-seeking two months after the end of the study provided a more valid measure of participants' behavior. For example, of the participants who were categorized as non-seekers at study-end, seven reported having made an appointment with a hearing professional at follow-up and 12 were planning to do so in the near future. Another limitation that affects the generalizability of the findings is the specific sample included in the study: older individuals living in Germany, using a smartphone in their daily life, mainly coming from big cities (57.3%) and having a monthly income above the national average net salary (€2,500) in 31.4% of cases. The conclusions about relevant personal factors for predicting hearing help-seeking may not be generalizable to people with hearing loss experiencing different socio-economic situations. For example, socio-economic factors were found to be the major limiting factor in seeking help for hearing difficulties in a South-African peri-urban community [79]. To address this socio-demographic limitation, future studies may consider alternative recruitment strategies to achieve a more diverse socio-demographic sample. These findings may also not generalize to different age groups. For example, when considering young adults with HL, other personal characteristics may be more important in predicting help-seeking.

Looking forward, the present study sets the milestones for the development and implementation of a short and concise profiling module in an existing or future mobile app for hearing screening, linked to targeted recommendations, complementing the audiological assessment in a modular environment. In the context of the present mobile study, individuals could provide any type of feedback in an open question format at the end of the study. Out of N = 185 individuals who completed the study, only one participant raised concerns related to usability and the user-interface. This participant suggested enhancing the contrast between the fonts and the background and increasing the size of the click buttons. As this feedback was provided towards the end of our data collection, we were unable to implement this suggestion in our design. Further studies that focus on user-interface and usability are necessary for the future implementation of such a module in mHealth solutions that target an older population should consider implementation strategies such as simplicity of design, naturalness of navigation and task flow, clear interface elements, feedback [80,81], large font sizes, contrasting colors, and clear, consistent and simple instructions [11]. Perceived ease of use and perceived usefulness should be targeted to promote acceptance and use of mHealth solutions [11].

Conclusions

This research provides initial knowledge towards a selection of tests and questionnaires that have been shown to predict hearing help-seeking in persons with self-reported hearing difficulties. From these, we derived conclusions for the implementation of an individual profiling algorithm in a hearing mHealth app. This study is innovative in that it considers a comprehensive range of personal characteristics and covariates previously cited in literature, including 25 assessment tools and 83 features, and narrows them down to identify a short selection of the most important predictors for profiling. Complementing the audiological assessments with such a profiling algorithm will enable an mHealth app to deliver targeted and efficient treatment recommendations depending on relevant individual characteristics. The benefits of such a profile module might also extend to other applications within a hearing mHealth app. Future studies might explore potential relationships between psychological traits and, among others, HA fitting preferences and endurance in the fine-tuning process towards an optimal aiding solution, openness to try new, elaborated technical solutions, or preference for particular app usability features. We have seen how predictive models that use machine learning algorithms can be used to explore complex association patterns of individual characteristics and behaviors, considering multiple predictors simultaneously and drawing robust conclusions through cross-validation approaches. This provides further evidence to the advance that the use of machine learning algorithms can bring to mHealth technology development [33]. mHealth solutions contribute to the evolving of hearing healthcare towards P4 medicine: predictive, preventive, personalized and participatory [75]. We have seen how individual profiling in a hearing mHealth app can identify non-seekers, acting as a preventive action to reduce the risk of a late intervention for HL. It can also provide clinicians with data-driven insights on the individual health profile of the user for tailored and personalized treatments. Moreover, it can enhance the empowerment and participation of the individual in their own hearing health care, promoting informed decision-making. Indeed, personalization strategies (such as tailored treatment recommendations) increase the effectiveness of mHealth interventions [82]. To conclude, a hearing mHealth app that provides targeted treatment recommendations could facilitate faster access to hearing-care services and subsequent earlier intervention, where needed, to pursue the long-term goal of achieving “hearing for all”.

Acknowledgements

This research has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2177/1 - Project ID 390895286. The study protocol was approved by the Research Ethics Committee of the Carl von Ossietzky Universität Oldenburg (08.09.2021, Drs.EK/2020/020-01). We thank all the participants who took part in this study and the Universities that helped us with recruitment: Universität Oldenburg, Gasthörstudium; Goethe-Universität Frankfurt am Main, Universität des 3. Lebensalters; Freie Universität Berlin, GasthörerCard Programm; University of Köln, Gasthörer- und Seniorenstudium; University of Kassel, Gasthörendenprogramm; University of Bielefeld, Studieren ab 50.

Data availability

Analyses scripts are available on Zenodo [83]. The data is available upon request to the corresponding authors. A preprint of the manuscript was published on medRxiv [84] in February 2023 and revised in August 2023. Preliminary results of this research were presented on symposia and conferences of the "Hearing4All" Cluster of Excellence, on the VCCA Conference 2022 and the DGPs Congress 2022.

Author’s contributions

GA, AH and MB conceptualized the study and were involved in protocol development, study design and data analysis. IK was involved in study design and data analysis. BK contributed to study conceptualization and obtained funding. GA was responsible for participants’ recruitment, data collection and wrote the first draft of the manuscript. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Abbreviations

CV: cross-validation
HA: hearing aids
HL: hearing loss
MCC: Matthews correlation coefficient
mHealth: mobile health
NA: not available
NB: naïve Bayes
RF: random forest
SNR: signal-to-noise ratio
SRT: speech recognition threshold
SVM: support vector machine

References

1. Haile LM, Kamenov K, Briant PS, et al. Hearing loss prevalence and years lived with disability, 1990–2019: findings from the Global Burden of Disease Study 2019. The Lancet Elsevier; 2021 Mar 13;397(10278):996–1009. PMID:33714390

2. Cox RM, Alexander GC, Gray GA. Who wants a hearing aid? Personality profiles of hearing aid seekers. Ear Hear 2005;26(1):12–26. PMID:15692301

3. Saunders GH, Chisolm TH, Wallhagen MI. Older adults and hearing help-seeking behaviors. Am J Audiol 2012;21(2):331–337. PMID:23233518

4. Sawyer CS, Armitage CJ, Munro KJ, Singh G, Dawes PD. Biopsychosocial Classification of Hearing Health Seeking in Adults Aged over 50 Years in England. Ear Hear 2020;1215–1225. PMID:31985532

5. Fisher DE, Li CM, Hoffman HJ, et al. Sex-specific predictors of hearing-aid use in older persons: The age, gene/environment susceptibility-Reykjavik study. Int J Audiol Taylor and Francis Ltd; 2015 Sep 2;54(9):634–641. PMID:25816699

6. Nixon G, Sarant J, Tomlin D, Dowell R. Hearing aid uptake, benefit, and use: The impact of hearing, cognition, and personal factors. J Speech Lang Hear Res 2021;64(2):651–663. PMID:33493400

7. Southall K, Gagné JP, Jennings MB. Stigma : A negative and a positive influence on help-seeking for adults with acquired hearing loss. Int J Audiol 2010;49:804–814. doi: 10.3109/14992027.2010.498447

8. Yong M, Willink A, McMahon C, McPherson B, Nieman CL, Reed NS, Lin FR. Access to adults' hearing aids: Policies and technologies used in eight countries. Bull World Health Organ World Health Organization; 2019;97(10):699–710. PMID:31656335

9. Simpson AN, Matthews LJ, Cassarly C, Dubno JR. Time from Hearing Aid Candidacy to Hearing Aid Adoption: A Longitudinal Cohort Study. Ear Hear Lippincott Williams and Wilkins; 2019 May 1;40(3):468–476. PMID:30085938

10. The Mobile Economy. Mob Econ. Available from: https://www.gsma.com/mobileeconomy

11. Hsieh KL, Fanning JT, Rogers WA, Wood TA, Sosnoff JJ. A Fall Risk mHealth App for Older Adults: Development and Usability Study. JMIR Aging 2018 Nov 20;1(2):e11569. doi: 10.2196/11569

12. Ahmad NA, Mat Ludin AF, Shahar S, Mohd Noah SA, Mohd Tohit N. Willingness, perceived barriers and motivators in adopting mobile applications for health-related interventions among older adults: a scoping review protocol. BMJ Open 2020 Mar;10(3):e033870. doi: 10.1136/bmjopen-2019-033870

13. Tong C, Kernoghan A, Lemmon K, Fernandes P, Elliott J, Sacco V, Bodemer S, Stolee P. Lessons and Reflections From an Extended Co-design Process Developing an mHealth App With and for Older Adults: Multiphase, Mixed Methods Study. JMIR Aging 2022 Oct 28;5(4):e39189. doi:

10.2196/39189

14. Masalski M, Grysiński T, Kręcicki T. Hearing Tests Based on Biologically Calibrated Mobile Devices: Comparison With Pure-Tone Audiometry. JMIR MHealth UHealth 2018 Jan 10;6(1):e7800. doi: 10.2196/mhealth.7800

15. Bright T, Pallawela D. Validated Smartphone-Based Apps for Ear and Hearing Assessments: A Review. JMIR Rehabil Assist Technol 2016 Dec 23;3(2):e6074. doi: 10.2196/rehab.6074

16. Li LYJ, Wang S-Y, Wu C-J, Tsai C-Y, Wu T-F, Lin Y-S. Screening for Hearing Impairment in Older Adults by Smartphone-Based Audiometry, Self-Perception, HHIE Screening Questionnaire, and Free-Field Voice Test: Comparative Evaluation of the Screening Accuracy With Standard Pure-Tone Audiometry. JMIR MHealth UHealth 2020 Oct 27;8(10):e17213. doi: 10.2196/17213

17. Lin H-YH, Chu Y-C, Lai Y-H, Cheng H-L, Lai F, Cheng Y-F, Liao W-H. A Smartphone-Based Approach to Screening for Sudden Sensorineural Hearing Loss: Cross-Sectional Validity Study. JMIR MHealth UHealth 2020 Nov 11;8(11):e23047. doi: 10.2196/23047

18. Chen C-H, Lin H-YH, Wang M-C, et al. Diagnostic Accuracy of Smartphone-Based Audiometry for Hearing Loss Detection: Meta-analysis. JMIR MHealth UHealth 2021 Sep 13;9(9):e28378. doi: 10.2196/28378

19. Paglialonga A, Cleveland Nielsen A, Ingo E, Barr C, Laplante-Lévesque A. eHealth and the hearing aid adult patient journey: a state-of-the-art review. Biomed Eng OnLine 2018 Jul 31;17:101. PMID:30064497

20. Frisby C, Eikelboom RH, Mahomed-Asmail F, et al. Mobile Health Hearing Aid Acclimatization and Support Program in Low-Income Communities: Feasibility Study. JMIR Form Res 2023 Aug 23;7(1):e46043. doi: 10.2196/46043

21. D'Onofrio KL, Zeng F-G. Tele-Audiology: Current State and Future Directions. Front Digit Health 2022;3. Available from: https://www.frontiersin.org/articles/10.3389/fdgth.2021.788103

22. Ratanjee-Vanmali H, Swanepoel DW, Laplante-Lévesque A. Patient Uptake, Experience, and Satisfaction Using Web-Based and Face-to-Face Hearing Health Services: Process Evaluation Study. J Med Internet Res 2020 Mar 20;22(3):e15875. doi: 10.2196/15875

23. Ratanjee-Vanmali H, Swanepoel DW, Laplante-Lévesque A. Characteristics, behaviours and readiness of persons seeking hearing healthcare online. Int J Audiol 2019;58(2):107–115. PMID:30289050

24. Frisby C, Eikelboom R, Mahomed-Asmail F, Kuper H, Swanepoel DW. m-Health Applications for Hearing Loss: A Scoping Review. Telemed E-Health Mary Ann Liebert, Inc., publishers; 2022 Aug;28(8):1090–1099. doi: 10.1089/tmj.2021.0460

25. Swanepoel DW. Advancing Equitable Hearing Care: Innovations in Technology and Service Delivery. Folia Phoniatr Logop 2023 Apr 14;75(4):201–207. doi: 10.1159/000530671

26. Knoetze M, Manchaiah V, Mothemela B, Swanepoel DW. Factors Influencing Hearing Help-Seeking and Hearing Aid Uptake in Adults: A Systematic Review of the Past Decade. Trends

Hear 2023;27:23312165231157256. PMID:36798964

27. Meyer C, Hickson L, Lovelock K, Lampert M, Khan A. An investigation of factors that influence help-seeking for hearing impairment in older adults. Int J Audiol 2014;53(S1):S3–S17. PMID:24405446

28. Jenstad L, Moon J. Systematic Review of Barriers and Facilitators to Hearing Aid Uptake in Older Adults. Audiol Res MDPI AG; 2011 Mar 23;1(1):91–96. doi: 10.4081/audiores.2011.e25

29. Vestergaard Knudsen L, Öberg M, Nielsen C, Naylor G, Kramer SE. Factors Influencing Help Seeking, Hearing Aid Uptake, Hearing Aid Use and Satisfaction With Hearing Aids: A Review of the Literature. Trends Amplif 2010;14(3):127–154. PMID:21109549

30. Schneider JM, Gopinath B, McMahon CM, Britt HC, Harrison CM, Usherwood T, Leeder SR, Mitchell P. Role of general practitioners in managing age-related hearing loss. Med J Aust John Wiley & Sons, Ltd; 2010 Jan 1;192(1):20–23. PMID:20047543

31. Convery E, Hickson L, Meyer C, Keidser G. Predictors of hearing loss self-management in older adults. Disabil Rehabil Informa UK Ltd.; 2019;41(17):2026–2035. PMID:29587551

32. Meyer C, Hickson L, Fletcher A. Identifying the barriers and facilitators to optimal hearing aid self-efficacy. Int J Audiol Informa Healthcare; 2014;53(S1). PMID:24447235

33. An Q, Rahman S, Zhou J, Kang JJ. A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges. Sensors 2023 Apr 22;23(9):4178. PMID:37177382

34. Rooksby J, Asadzadeh P, Morrison A, McCallum C, Gray C, Chalmers M. Implementing ethics for a mobile app deployment. Proc 28th Aust Conf Comput-Hum Interact - OzCHI 16 Launceston, Tasmania, Australia: ACM Press; 2016. p. 406–415. doi: 10.1145/3010915.3010919

35. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Privacy and Security in Mobile Health Apps: A Review and Recommendations. J Med Syst 2014 Dec 7;39(1):181. doi: 10.1007/s10916-014-0181-3

36. Arslan RC, Walther MP, Tata CS. formr: A study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using R. Behav Res Methods 2019 521 Springer; 2019 Apr 1;52(1):376–387. doi: 10.3758/S13428-019-01236-Y

37. https://www.twilio.com/. Available from: https://www.twilio.com/

38. Senese VP, Ruotolo F, Ruggiero G, Iachini T. The italian version of the weinstein noise sensitivity scale: Measurement invariance across age, sex, and context. Eur J Psychol Assess 2012;28(2):118–124. doi: 10.1027/1015-5759/a000099

39. Meis M, Huber R, Fischer RL, Schulte M, Spilski J, Meister H. Development and psychometric properties of the sound preference and hearing habits questionnaire (SP-HHQ). Int J Audiol Informa UK Limited, trading as Taylor & Francis Group; 2018;57(0):S118–S129. PMID:27875658

40. Goldberg LR. An alternative "description of personality": The Big-Five factor structure. J Pers Soc Psychol US: American Psychological Association; 1990;59:1216–1229. doi: 10.1037/0022-

3514.59.6.1216

41. Aron EN, Aron A. Sensory-Processing Sensitivity and Its Relation to Introversion and Emotionality. J Pers Soc Psychol American Psychological Association Inc.; 1997;73(2):345–368. PMID:9248053

42. Posner J, Russell JA, Peterson BS. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev Psychopathol NIH Public Access; 2005 Jul;17(3):715. PMID:16262989

43. Scott SB, Sliwinski MJ, Zawadzki M, Stawski RS, Kim J, Marcusson-Clavertz D, Lanza ST, Conroy DE, Buxton O, Almeida DM, Smyth JM. A Coordinated Analysis of Variance in Affect in Daily Life. Assessment SAGE Publications Inc.; 2020 Dec 1;27(8):1683–1698. PMID:30198310

44. Smits C, Kapteyn T, Houtgast T. Development and validation of an automatic speech-in-noise screening test by telephone. Int J Audiol 2004 Feb 1;43:15–28. doi: 10.1080/14992020400050004

45. Buschermöhle M, Wagener KC, Berg D, Meis M, Kollmeier B. The German Digits Triplets Test (Part I): Implementations for Telephone, Internet and Mobile Devices. Audiol Acoust Für Audiol 2014;54(4):139–145.

46. Van den Borre E, Denys S, van Wieringen A, Wouters J. The digit triplet test: a scoping review. Int J Audiol Taylor & Francis; 2021 Dec 1;60(12):946–963. PMID:33840339

47. Denys S, Hofmann M, van Wieringen A, Wouters J. Improving the efficiency of the digit triplet test using digit scoring with variable adaptive step sizes. Int J Audiol Taylor & Francis; 2019 Oct 3;58(10):670–677. PMID:31187664

48. Potgieter J-M, Swanepoel DW, Myburgh HC, Hopper TC, Smits C. Development and validation of a smartphone-based digits-in-noise hearing test in South African English. Int J Audiol Taylor & Francis; 2016 Jul 2;55(7):405–411. PMID:27121117

49. Smits C, Merkus P, Houtgast T. How we do it: The Dutch functional hearing-screening tests by telephone and internet. Clin Otolaryngol 2006 Oct;31(5):436–440. PMID:17014457

50. Löw T, Bach J-H, Buschermöhle M, Bisitz T. Development of new procedures for the Digit Triplets Test as a reliable screening tool. 21 Jahrestag Dtsch Ges Für Audiol 2018. Available from: https://www.dga-ev.com/fileadmin/dga2018/site/data/final/0146.pdf

51. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2019. Available from: https://www.r-project.org/

52. Arslan RC. Formr.Org Survey Framework Utility R Package. Zenodo; 2018. doi: 10.5281/ZENODO.594253

53. Bischl B, Lang M, Kotthoff L, Schiffner J, Richter J, Studerus E, Casalicchio G, Jones ZM. mlr: Machine Learning in R. J Mach Learn Res 2017;17(170):1–5.

54. Rhys HI. Machine Learning with R, the tidyverse, and mlr. Manning Publications Co.; 2020. ISBN:978-1-61729-657-4

55. Becker M, Binder M, Bischl B, Foss N, Kotthoff L, Lang M. mlr3book. 2022. Available from: https://mlr3book.mlr-org.com

56. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning. Design. New York, NY: Springer New York; 2013. PMID:10911016 ISBN:978-1-4614-7137-0

57. Probst P, Bischl B. Tunability: Importance of Hyperparameters of Machine Learning Algorithms. J Mach Learn Res. 2019 p. 1–32. Available from: http://jmlr.org/papers/v20/18-444.html.

58. Becker M, Ullmann T, Lang M, Bischl B, Richter J, Binder M. mlr-org - Practical Tuning Series - Tune a Support Vector Machine. 2021. Available from: https://mlr-org.com/gallery/2021-03-09-practical-tuning-series-tune-a-support-vector-machine/#tuning-search-space

59. Hsu C-W, Chang C-C, Lin C-J. A Practical Guide to Support Vector Classification. 2016; Available from: http://www.csie.ntu.edu.tw/~cjlin

60. Ali A, Mariyam Shamsuddin S, Ralescu A, Ralescu AL. Classification with class imbalance problem: A review. Classif Int J Adv Soft Compu Appl 2013;5(3). Available from: https://www.researchgate.net/publication/288228469

61. Delgado R, Tibau X-A. Why Cohen's Kappa should be avoided as performance measure in classification. PLOS ONE Public Library of Science; 2019 set;14(9):e0222916. doi: 10.1371/journal.pone.0222916

62. Kuhn M, Vaughan D, Hvitfeldt E, RStudio. yardstick: Tidy Characterizations of Model Performance. 2022. Available from: https://CRAN.R-project.org/package=yardstick

63. Jurman G, Riccadonna S, Furlanello C. A Comparison of MCC and CEN Error Measures in Multi-Class Prediction. PLOS ONE Public Library of Science; 2012 ago;7(8):e41882. doi: 10.1371/journal.pone.0041882

64. Saarela M, Jauhiainen S. Comparison of feature importance measures as explanations for classification models. SN Appl Sci 2021;3(2). doi: 10.1007/s42452-021-04148-9

65. Kuhn M. The caret Package. 2019. Available from: https://topepo.github.io/caret/index.html

66. Kuhn M. caret: Classification and Regression Training. 2020. Available from: https://CRAN.R-project.org/package=caret

67. Kiessling J, Grugel L, Meister H, Meis M. Übertragung der Fragebögen SADL , ECHO und SSQ ins Deutsche und deren Evaluation. Z Für Audiol 2011;50(1):6–16.

68. Saunders GH, Cienkowski KM, Forsline A, Fausti S. Normative data for the Attitudes towards Loss of Hearing Questionnaire. J Am Acad Audiol 2005;16(9):637–652. PMID:16515136

69. Nuesse, Schlueter A, Lemke U, Holube I. Self-reported hearing handicap in adults aged 55 to 81 years is modulated by hearing abilities, frailty, mental health, and willingness to use hearing aids. Int J Audiol Taylor and Francis Ltd.; 2021;60(S2):71–79. PMID:33459099

70. Konrad S, Herzberg PY. Psychometric properties and validation of a German high Sensitive person scale (HSPS-G). Eur J Psychol Assess 2019;35(3):364–378. doi: 10.1027/1015-5759/a000411

71. Drixler K, Morfeld M, Glaesmer H, Brähler E, Wirtz MA. Validierung der Messung gesundheitsbezogener Lebensqualität mittels des Short-Form-Health-Survey-12 (SF-12 Version 2.0) in einer deutschen Normstichprobe. Z Psychosom Med Psychother 2020;66(3):272–286. PMID:32876561

72. Borkenau P, Ostendorf F. NEO-Fünf-Faktoren-Inventar (NEO-FFI) nach Costa und McCrae: Handanweisung. Göttingen: Hogrefe.; 1993. doi: 10.1026//0012-1924.48.1.19

73. Kuhlmann I, Angonese G, Thiel C, Kollmeier B, Hildebrandt A. Are there good days and bad days for hearing? Quantifying day-to-day intraindividual speech perception variability in older and younger adults. J Exp Psychol Hum Percept Perform US: American Psychological Association; 2023;49(11):1377–1394. doi: 10.1037/xhp0001159

74. Ramos MD. Exploring the relationship between planned behavior and self-determination theory on health-seeking behavior among older adults with hearing impairment. Geriatr Nur (Lond) 2023 Jul 1;52:1–7. doi: 10.1016/j.gerinurse.2023.05.001

75. Flores M, Glusman G, Brogaard K, Price ND, Hood L. P4 medicine: how systems medicine will transform the healthcare sector and society. Pers Med 2013;10(6):565–576. PMID:25342952

76. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007;133(4):673–693. doi: 10.1037/0033-2909.133.4.673

77. Hirsh JB, Kang SK, Bodenhausen GV. Personalized Persuasion: Tailoring Persuasive Appeals to Recipients' Personality Traits. Psychol Sci SAGE Publications Inc; 2012 Jun 1;23(6):578–581. doi: 10.1177/0956797611436349

78. Ferguson E. Personality is of central concern to understand health: towards a theoretical model for health psychology. Health Psychol Rev Routledge; 2013 May 1;7(sup1):S32–S70. PMID:23772230

79. Mtimkulu TK, Khoza-Shangase K, Petrocchi-Bartal L. Barriers and facilitators influencing hearing help-seeking behaviors for adults in a peri-urban community in South Africa: a preventive audiology study. Front Public Health 2023;11. PMID: 37920587

80. Wildenbos GA, Peute LW, Jaspers MWM. A framework for evaluating mHealth tools for Older Patients on Usability. Stud Health Technol Inform 2015;210:783–787. PMID:25991261

81. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: A literature based framework (MOLD-US). Int J Med Inf 2018 Jun 1;114:66–75. doi: 10.1016/j.ijmedinf.2018.03.012

82. Rivera-Romero O, Gabarron E, Ropero J, Denecke K. Designing personalised mHealth solutions: An overview. J Biomed Inform 2023 Oct 1;146:104500. doi: 10.1016/j.jbi.2023.104500

83. Angonese G. Machine learning and feature importance for Predicting hearing help-seeking. Zenodo; 2023. doi: 10.5281/zenodo.7635920

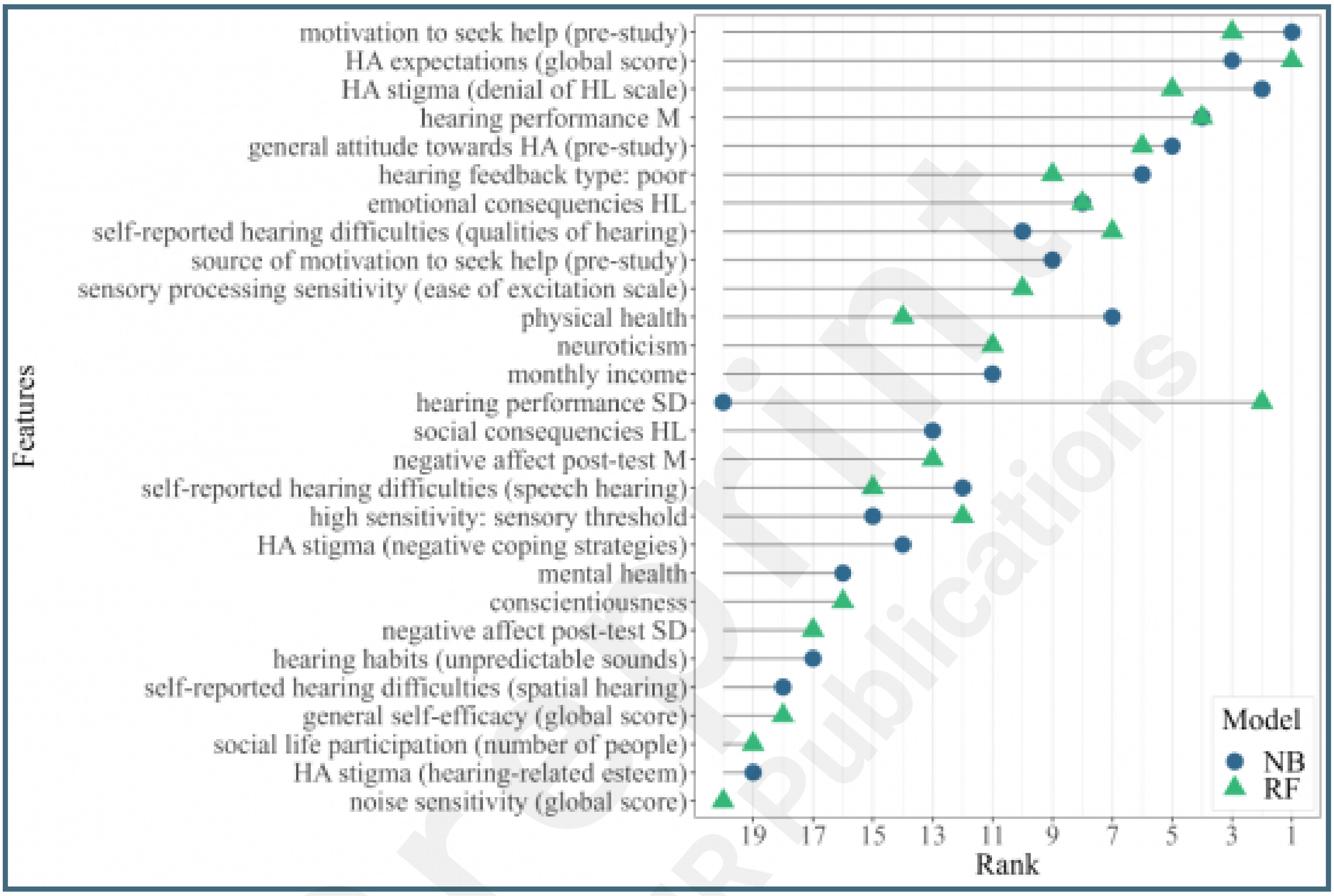
84. Giulia Angonese, Mareike Buhl, Inka Kuhlmann, Birger Kollmeier, Andrea Hildebrandt.

Predicting hearing help-seeking: What features are important for a profiling module within a hearing mHealth application? medRxiv 2023 Jan 1;2023.02.14.23285904. doi: 10.1101/2023.02.14.23285904

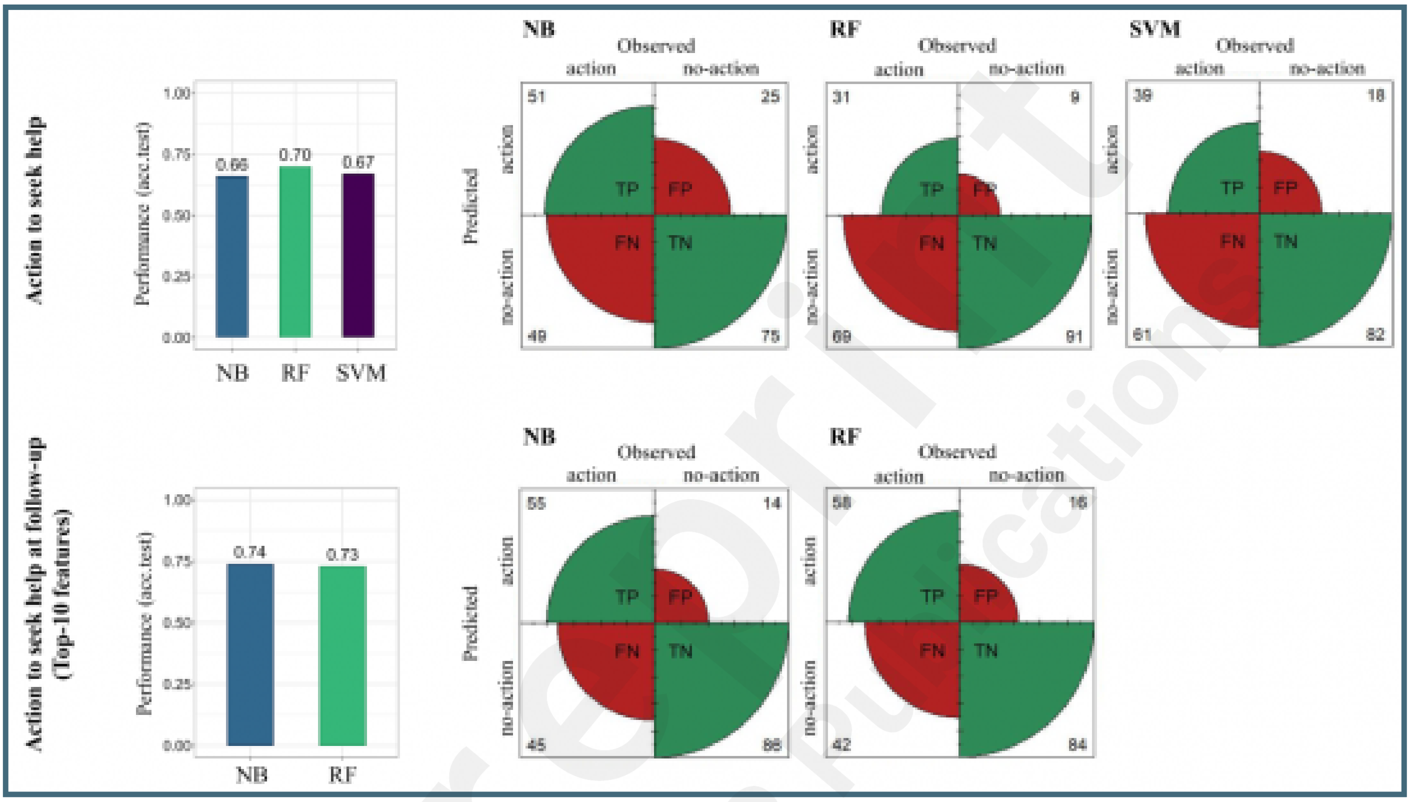
Supplementary Files

Figures

Ranking of the most important features to predict action to seek help at study-end. Importance ranking are shown for the 20 most important features for the two models (NB and RF) trained on the full dataset (N1=185), as summarized in Table 5. A total of 28 features are arranged on the y-axis with respect to their average ranking between the two models. (HL: hearing loss; HA: hearing aids; M: mean; SD: standard deviation; NB: Naïve Bayes; RF: Random Forest).



Visualization of the model-specific performances (left) and relative confusion matrices (right) for the outcomes action to seek help (first row) measured at study end (N1 = 185) and action to seek help at follow-up (second row) considering the set of top-10 most important features (N2 = 131). The bar plots on the left show the model-specific performance for each outcome measured as prediction accuracy on the test set. Each confusion matrix plot illustrates the class-specific relative frequencies for each observed and predicted class, and provides a more immediate overview of each model's sensitivity (TP) and specificity (TN). (TP: True positive; TN: True Negative; FP: False Positive; FN: False Negative; NB: Naïve Bayes classifier; RF: Random Forest; SVM: Support Vector Machine).



Multimedia Appendixes

Study design overview.
URL: http://asset.jmir.pub/assets/74b4f4c9ff57f35b3551e903fc437239.pdf

Hearing test feedback.
URL: http://asset.jmir.pub/assets/18397fd1ef078e0dca0c23e8cb5305f0.pdf

Feature importance values.
URL: http://asset.jmir.pub/assets/94299c2b59ddef638e2b3b1d5babd3b7.pdf

Features correlation plots.
URL: http://asset.jmir.pub/assets/68475b06dda37e96527795f42c9141eb.pdf