

Northumbria Research Link

Citation: Anwar, Naveed and Oakes, Michael (2013) Decision support system for the selection of an ITE or a BTE hearing aid. International Journal of Computer Applications, 76 (14). pp. 37-42. ISSN 0975-8887

Published by: Inderscience

URL: <http://dx.doi.org/10.5120/13318-0936> <<http://dx.doi.org/10.5120/13318-0936>>

This version was downloaded from Northumbria Research Link:
<https://nrl.northumbria.ac.uk/id/eprint/22200/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

Decision Support System for the Selection of an ITE or a BTE Hearing Aid

Muhammad Naveed Anwar
Knowledge Media Institute
The Open University
Milton Keynes, UK

Michael Philip Oakes
Faculty of Applied Sciences
University of Sunderland Sunderland, UK

ABSTRACT

The purpose of this research is to mine a large set of heterogeneous audiology data to create a decision support system (DSS) to choose between two hearing aid types (ITE and BTE aid). This research is based on the data analysis of audiology data using various statistical and data mining techniques. It uses the data of a large NHS (National Health Services, UK) facility. It uses 180,000 records (covering more than 23,000 different patients) from a hearing aid clinic. The developed system uses an unconventional method to predict hearing aid type for a patient and it can be used as a second opinion by audiologists for complex cases. After modifying the system to take account of the feedback from a professional audiologist, the success rates obtained were in the ranges 63 to 66 percent. In this research an automatic system was developed to choose between an ITE or a BTE hearing aid type with an explanation facility that can be used as a second opinion by audiologist in cases where the choice of an ITE or a BTE hearing aid is not clear cut. This analysis of audiology data and DSS will provide supplementary information for audiology experts and hearing aid dispensers. This type of system may also be of interest to manufacturers of hearing technologies in using as a ready means for their telephone customer services staff to check data, discovering data in audiology records will also be good for general awareness about the suitability of hearing aid type.

General Terms

Data mining, data analysis, medical records processing.

Keywords

ITE (in the ear) hearing aid; BTE (behind the ear) hearing aid; decision support system; Naïve Bayesian analysis; F-score.

1. INTRODUCTION

There is tremendous growth in the amount of data produced in the medical domain [1] and each year the amount of data increases and new terms emerge. The relating of medical terms, diseases and other features to find their causes is important for taking decisions and making predictions. Much work in terms of mining, retrieval of information and finding associations among diseases and other related attributes has been done in the medical domain. Many approaches, including statistical and neural approaches have been proposed for medical data mining which produce information that helps in problem solving and taking decisions [2-4].

Over nine million people in the UK (or 1 in 7 of the population) suffer from hearing loss [5], among whom the consequences of hearing loss are varied. People with hearing loss are found to face the following problems: (1) serious

depression due to loss of hearing, (2) problems in speech and communication, and (3) avoiding public places due to unsafe environments like heavy traffic.

In spite of the above, it has been found that only 85% of people use their hearing aid(s) regularly, 12% only occasionally and 3% never, as they may or may not be happy with their hearing aids [6]. Therefore, motivated by the need to prescribe the right hearing aid to make a patient satisfied, there is a need for a decision support system (DSS) in audiology. Thus, this research employs a set of audiology data taken first hand by hospital staff. The research involves mining heterogeneous audiology data to get useful information that could not be obtained otherwise, enabling the findings of many possible associations in this large amount of data. The findings will be helpful in the decision to prescribe BTE (behind the ear) or ITE (in the ear) hearing aids. Although ITE hearing aids are not generally available on the NHS - National Health Service since they are more expensive than BTE aids [7], the data used in this research comes from a large NHS audiology facility, which does offer ITE hearing aids on the NHS. This adds an important feature to this data set.

In the medical domain, in particular in the audiology domain, there is a substantial amount of heterogeneous data available. This research deals with three different types of data: (1) audiograms (graphs of hearing ability at different frequencies), e.g. 65, 65, 35, 40, 45, 55, 0, 10, 25, 40, 50. The first six values are AC thresholds (the faintest sound that the patient can hear in decibels) at 250, 500, 1000, 2000, 4000 and 8000 Hz, and the following five values are the BC thresholds for the same set of frequencies except 8000 Hz, (2) structured tabular data (such as, gender, diagnosis, hearing aid type.), e.g., M, TINNITUS, ITENN, and (3) unstructured text (specific observations made about each patient's case, referred to as free text, e.g., AT REV LT ITENL TO ITENN AS INSUFFICIENT GAIN-SOUNDED MUCH BETTER!, which is shorthand for "At review, the left ITENL hearing aid was replaced by an ITENN hearing aid, as the old one had insufficient gain. The new one sounded much better").

The originality of the work is: (1) this research involves mining heterogeneous audiology data set (numeric, categorical and text), and so the work is useful for medical mining in general, as medical records tend to be of all three data types, (2) this research makes use of a very large data set compared with others as it contains 180,000 individual records covering more than 23,000 different patients from a hearing aid clinic and this is the first hand data collected by audiology technicians, (3) a unique data set related to ITE/BTE hearing aids is used, as this large NHS audiology facility is one of those hospitals who dispense both BTE and

ITE hearing aids, (4) an audiologist from a large NHS facility has reflected on and given feedback on the work, which gives unique insights into the findings and decision making processes, (5) this research considers many predictive factors together in a single model, which has not been performed in previous studies, and (6) this research will improve the quality of service in audiology clinics, at affordable cost, by prescribing for patients more accurately.

In this research, a data mining approach is used to discover the factors contributing to the choice between ITE and BTE hearing aids. Previous research in these areas has mainly used questionnaires or clinical trials with smaller datasets [2, 4, 8-10].

2. OBJECTIVE

This research amalgamates the heterogeneous audiology data into a form that is useful for building statistical models. Candidate factors such as air and bone conduction frequencies, age, gender, diagnosis, masker, mould and free text words associated with the patients are used as inputs and the DSS gives as the output a decision as to whether the patient would be more likely to prefer an ITE or a BTE hearing aid type.

3. MATERIALS AND METHODS

3.1 Methodology

Starting with the heterogeneous set of variables in the audiology data, the authors identified those which were mentioned in the literature as having a bearing on the choice of hearing aid type. Of these, just those were retained that were found to be related with the research question in a series of pilot studies involving statistical and data mining techniques [11-12]. In order to combine the candidate factors for the research question into a single model, the technique of Naïve Bayesian analysis is used. The Naïve Bayesian analysis can only represent simple distributions and give probabilistic estimations rather than predictions [13-14]. It allows predictions to be ranked, and their expected cost to be minimised. The advantage of this technique for the combination of evidence is that it is easy to see which variables contributed to the final decision.

3.2 Validation of Technique

For the evaluation, the results computed by the final models were compared with real hearing scenarios. Below are the validation methods used in this research.

- To validate the overall results of Naïve Bayesian analysis, a five-fold cross validation was used. 80% of the data sample was used for constructing the model and the remaining 20% of the data sample was used for testing this model, which is a common machine learning approach to train and test a classifier. Each record of the testing data was used as follows: the values of the predictor variables were fed into the final model, and the output of the model (hearing aid type or presence of a masker) was compared with the value in the corresponding field of the test data record. Precision, recall and the F-measure are standard information retrieval metrics which were used for comparing predicted and “correct” outputs [15]. Precision is the ratio of cases where the model and the “correct” output agree to the total number of times a particular decision was made by the model, and recall is the ratio of agreements to the number of times that

decision should have been made by the model. The F-measure combines precision and recall into a single measure.

- The data mining results were compared with the clinical experience of a professional audiologist.

3.3 Deliverables

The final models have been implemented in a spreadsheet, because this can hold both data and processing together. A spreadsheet can conveniently allow “tracing back” to show how the final model’s decision was arrived at. Another reason for using the spreadsheet is that it can be operated by audiology professionals, and reduce the time required to learn to use the software.

3.4 Assumption

It is assumed that in every record in the audiology data set the audiologists’ original decisions are correct, for example, in the choice of a hearing aid of type (ITE or BTE).

3.5 Limitation of the study

The limitation of the study is that this audiology data was collected between 1992 and 2001, but more recent data could not be provided. The other factors mentioned in the literature which could not be tested with this data were the greater cosmetic acceptability of ITE aids, comfort in wearing of aids, ease of use with spectacles and sound quality.

4. EXPERIMENTS

Naïve Bayes Analysis is a machine learning algorithm, and is a special case of a Bayesian network. It is based on Bayes’ rule and multiplies probabilities when events or attributes are independent [16]. There are three elements involved with Naïve Bayesian analysis: prior probability, likelihood ratio and posterior probability. Prior probability is the initial probability of being true of any hypothesis before an experiment is performed. Likelihood is the conditional probability of the information, provided that the hypothesis is true. Posterior probability is the conditional probability of a hypothesis being true given the value of the observed information. Each piece of evidence available is considered in turn.

The Naïve Bayesian analysis was performed on the records which had all fields filled for the right ear: AC (air conduction) and BC (bone conduction) thresholds, gender, age and comments text keywords (5,736 records), of which 128 also had non-null entries for diagnosis, 101 had non-null entries for masker, and 3983 had non-null entries for mould.

The prior odds for our calculation was the relative likelihoods of preferring one type of hearing aid before any of the individual characteristics were considered. The prior odds were calculated by using the total number of records of ITE and BTE, which in this case are 3073 and 2663 respectively (given in Table 1), and the total number of records, which are 5736. Now, prior odds are calculated by using Eq. 1:

$$\text{Prior odds} = \text{Prior} / (1 - \text{Prior}) \quad [\text{Eq. 1}]$$

Where $\text{Prior} = \frac{\text{Total_number_of_records_of_ITE}}{\text{Total_number_of_records}} = \frac{3073}{5736} = 0.5357$, and $(1 - \text{Prior}) = 1 - 0.5357 = 0.4643$, so the prior odds for ITE and BTE were 1.15 and 0.87 respectively. This means that patients will be slightly more likely to require an ITE aid, and slightly less likely to require a BTE aid. In an individual audiology record, the AC threshold at 250 Hz (that is, AC250) might be 30dB, then Laplace smoothing was applied to the

observed number of subjects with this threshold, that is 1 was added to all observed values, so that none of the values were left equal to zero. So, the observed values, that is the frequencies of ITE and BTE become 1735 and 842 respectively and the total of ITE and BTE frequencies become 3077 and 2667 respectively. Then, the likelihood ratios are calculated by

$$Likelihood_ratios_ITE_at_AC250 = \frac{(Frequency_of_ITE/Total_ITE_frequencies)}{(Frequency_of_BTE/Total_BTE_frequencies)} \quad [Eq. 2]$$

$$Likelihood_ratios_ITE_at_AC250 = \frac{(1735/3077)}{(842/2667)} = 1.79 \quad [Eq. 3]$$

$$Likelihood_ratios_BTE_at_AC250 = \frac{(Frequency_of_BTE/Total_BTE_frequencies)}{(Frequency_of_ITE/Total_ITE_frequencies)} \quad [Eq. 4]$$

$$Likelihood_ratios_BTE_at_AC250 = \frac{(842/2667)}{(1735/3077)} = 0.56 \quad [Eq. 5]$$

Table 1. AC250 Hz categories with observed frequencies and likelihood ratios of ITE and BTE aids.

AC250	ITE	BTE	Row total	Likelihood ratios	
				ITE	BTE
≤40	1734 (67%)	841 (33%)	2575 (45%)	1.79	0.56
≤55	837 (55%)	697 (45%)	1534 (27%)	1	0.96
≤75	407 (36%)	719 (64%)	1126 (20%)	0.5	2
>75	95 (19%)	406 (81%)	501 (9%)	0.2	5
Column total	3073 (54%)	2663 (46%)	5736 (100%)		

In Eq. 3 and Eq. 5, the likelihood ratios of 1.79 (ITE) and 0.56 (BTE) are for AC250 only, suggesting that the patient will be slightly more likely to require an ITE aid, and slightly less likely to require a BTE aid. When the next field of the record (in this case the AC threshold at 500Hz) is taken into account, the process is repeated. So, once the likelihood ratios for all the record fields in Table 2 have been found, the final posterior odds are obtained by taking the product of the prior odds and all the discovered likelihood ratios, which are respectively 41.80 and 0.02 for ITE and BTE, suggesting that it is much more likely that the patient would prefer an ITE hearing aid type.

Testing of these tables models showed that overall there was 80.25% agreement between the predictions of this model and the actual hearing aid chosen by the audiologist (as given in the “type” field). The agreement rate was higher for patients fitted with BTE aids (85.71%) than for those fitted with ITE aids (75.70%), as shown in Table 3. The results were analyzed according to the precision, recall and F-measures [15] using Eq. 6 to 8 respectively. In the Naïve Bayesian analysis model all the candidate variables (AC and BC frequencies, age, gender, diagnosis, tinnitus maskers, mould, and free text keywords) described earlier are included. In Table 3, 782 and 651 are the counts of ITE and BTE aids respectively in the

human-annotated test data, while 685 and 748 are the counts of ITE and BTE aids respectively in the machine predicted results. The overall agreement is much better than random (50%). In this case, the ZeroR baseline would be to assume that all the patients should be assigned the more commonly prescribed type of hearing aid. In this test set 782 out of 1433 patients in the test set were given ITE aids, so simply assigning all the patients this type of aid would provide 54.6% agreement. The agreements found for ITE and BTE are higher than this.

$$P = \frac{Agreements_of_machine_and_human}{Total_number_in_that_category_by_machine} \Rightarrow P_{ITE} = \frac{592}{(592+93)} = 0.86 \quad [Eq. 6]$$

$$R = \frac{Agreements_of_machine_and_human}{Total_no._in_that_category_in_reality} \Rightarrow R_{ITE} = \frac{592}{(592+190)} = 0.76 \quad [Eq. 7]$$

$$F_{ITE} = 2 \times \frac{P \times R}{P + R} = 2 \times \frac{0.86 \times 0.76}{0.86 + 0.76} = 0.81 \quad [Eq. 8]$$

Table 2. Evidence in an individual patient record combined to obtain the posterior odds of a patient requiring an ITE or a BTE aid.

Candidate variables (Database record)	Actual values	Likelihood ratios	
		ITE	BTE
Age	77	1.22	0.82
Gender	Male	1.09	0.92
AC250	30	1.79	0.56
AC500	35	1.91	0.52
AC1000	45	1.06	0.94
AC2000	50	1.33	0.75
AC4000	70	1.15	0.87
AC8000	65	1.34	0.75
BC250	25	1.16	0.86
BC500	35	1.14	0.88
BC1000	35	1.12	0.89
BC2000	50	1.12	0.89
BC4000	65	0.87	1.15
Diagnosis	Tinnitus	1.26	0.80
Hearing aid type	ITE	To be found	
Masker	Yes	1.83	0.55
Mould	No mould given	1	1
Free text keywords	Rev	1.11	0.9
Posterior odds		41.80	0.02

Table 3. ITE/BTE aid predicted results for Naïve Bayesian analysis.

Machine results (Naïve Bayesian analysis)	Human (actual data)		
	ITE	BTE	Total
ITE	592 (75.70%)	190 (24.30%)	782
BTE	93 (14.29%)	558 (85.71%)	651
Total	685	748	1433

5. VALIDATION OF RESULTS

This section includes the validation methods by creating different sample splits of the audiology data for Naïve Bayesian analysis, and by taking the comments from an audiology expert on the data tables and results.

5.1 Splitting of audiology data and comparison of different splits for ITE/BTE aid

Before performing the Naïve Bayesian analysis, all (11,068) records were divided into two parts, one containing 80 percent (8,855) of the records and the other containing the remaining 20 percent (2,213) of these records. The 80 percent subset was used as the training set for model construction, and the remaining records for testing the model. For the validation of results, five-fold cross validation (repeated sub-sampling of the data to produce five non-overlapping test sets for an unbiased estimation of model accuracy) was performed. The models constructed were evaluated for correctly classifying the records using the percentage accuracy, precision, recall and F-measure.

With Naïve Bayesian analysis, the similarity was in the range 80 to 81%, precision was in the ranges 0.86 to 0.88 for ITE and 0.74 to 0.76 for BTE, recall was 0.74 to 0.76 for ITE and 0.86 to 0.88 for BTE, and the F-measure was 0.80 to 0.81 for ITE and 0.80 to 0.82 for BTE. These results show that for each run, both the final model and the success rates were similar for all five cross validations.

5.2 Audiology expert comments

As part of the validation process, the data tables behind the models were shown to an audiologist for comments and feedback, which are given below for different associated factors.

5.2.1 Audiograms

The audiologist stated that all patients irrespective of hearing loss are suitable for a BTE aid because its range of performances covers all levels of hearing loss. ITE aids have an upper limit of about 75dB (this is a “rule of thumb”) but BTE can tolerate more at high frequencies. ITE aids are more susceptible to the occlusion effect (feedback), since the microphone is close to the output. Thus, there is a limit for ITE aid feedback at high frequencies, so an ITE aid is not given if the high frequency loss is greater than 75dB. Also, conductive hearing losses tend to be flatter, and high frequency losses tend to be sensorineural, so the effect of frequency in the audiology data is related to whether the patients has a conductive or SNHL. Thus, ITE aids are less suitable for conductive hearing losses.

The contingency tables derived from the raw data from which the probabilities in the models were calculated relating hearing loss at to AC250, AC1000 and AC4000 to the choice of ITE/BTE aids are given in Table 1, Table 4 and Table 5 respectively. In Table 1, the highest percentage of patients given an ITE aid was 67% for AC250 at “<=40”, and the highest percentage of patients given a BTE aid was 81% for AC250 “>75”. Also, for AC250 at “>75” likelihood ratios obtained using Naïve Bayesian analysis for ITE and BTE aids were approximately 0.2 and 5 respectively, which shows that more patients were assigned BTE aids. Similarly, in Table 4 and Table 5, the highest percentage of patients given an ITE aid were 68% and 63% with likelihood ratios 1.88 and 1.45 for AC1000 at “<=40” and AC4000 at “<=55” respectively, and for a BTE aid were 89% and 60% with likelihood ratios

9.07 and 0.69 for AC1000 at “>75” and AC4000 at “>75” respectively. This shows that for low frequency hearing losses, ITE hearing aids and for high frequency hearing losses, BTE hearing aids are prescribed, which is consistent with the audiologist’s comments given above.

Table 4. AC1000 Hz categories with observed frequencies and likelihood ratios of ITE and BTE aids.

AC1000	ITE	BTE	Row total	Likelihood ratios	
				ITE	BTE
<=40	1529 (68%)	704 (32%)	2233 (39%)	1.88	0.53
<=55	1027 (55%)	836 (45%)	1863 (32%)	1.06	0.94
<=75	474 (38%)	778 (62%)	1252 (22%)	0.53	1.89
>75	43 (11%)	345 (89%)	388 (7%)	0.11	9.07
Column total	3073 (54%)	2663 (46%)	5736 (100%)		

Table 5. AC4000 Hz categories with observed frequencies and likelihood ratios of ITE and BTE aids.

The two main predictors of hearing aid type are air conduction thresholds and air-bone gap (difference between BC and AC thresholds). There are computer packages for the fitting of hearing aids and selection of the aid to be prescribed. In the NOAH package, the amplification required depends on the overall hearing loss, with a correction factor for the degree of conductive loss, so a future version of the DSS developed in this research should use as inputs AC and air-bone gap (BC-AC), rather than BC at each frequency. The audiologist also made the suggestion that in future, null inputs should be

AC4000	ITE	BTE	Row total	Likelihood ratios	
				ITE	BTE
<=40	289 (62%)	174 (38%)	463 (8%)	1.44	0.70
<=55	672 (63%)	402 (37%)	1074 (19%)	1.45	0.69
<=75	1445 (57%)	1092 (43%)	2537 (44%)	1.15	0.87
>75	667 (40%)	995 (60%)	1662 (29%)	0.58	1.72
Column total	3073 (54%)	2663 (46%)	5736 (100%)		

allowed for BC at 250 Hz, as this is no longer routinely measured (audiometers are no longer calibrated for this).

5.2.2 Age

In Table 6, which cross tabulates age vs. ITE/BTE aids, ITE aids were given relatively often to patients in the 60-78 age group, and patients more than 78 years of age were more often given BTE aids. Similarly, using Naïve Bayesian analysis the likelihood ratios for ITE and BTE aids in Table 6 for Age at “>78” were approximately 0.82 and 1.23 respectively, suggesting a BTE aid.

The audiologist commented that patients over the age of 78 prefer BTE aids, as this is the cohort with the greatest degree of hearing loss, since hearing loss advances with age. Thus, a

higher proportion of patients over the age of 78 are ineligible for ITE aids and have not been offered one, as shown in Table 6. Patients under the age of 60 might include children, who are mostly not given ITE aids. Some of these children will have profound hereditary hearing loss, so will need the most powerful aids available. Also, most other children will be excluded from ITE aids, as their classrooms have an interface between a radio link and hearing aid (the teacher wears a transmitter) which is not available in ITE. Thus age 0 to 18 is an almost exclusively BTE cohort. In the 20 – 60 age group, they will not yet have age-induced hearing loss (AIHL), so someone in this group has either had their hearing loss since childhood or it has been acquired later, usually due to disease (generally causing conductive hearing losses such as eardrum damage or otosclerosis). For manual dexterity ITE aids are easier to use, that is, they have simpler controls and are easier to access, which helps older people, accounting for their popularity in the 70 to 78 age group.

Table 6. Age categories with observed frequencies and likelihood ratios of ITE and BTE aids.

Age	ITE	BTE	Row total	Likelihood ratios	
				ITE	BTE
<=60	729 (52%)	673 (48%)	1402 (24%)	0.94	1.07
<=70	779 (55%)	631 (45%)	1410 (25%)	1.07	0.93
<=78	864 (58%)	614 (42%)	1478 (26%)	1.22	0.82
>78	701 (48%)	745 (52%)	1446 (25%)	0.82	1.23
Column total	3073 (54%)	2663 (46%)	5736 (100%)		

5.2.3 Sex

In the data table showing the relation between sex and ITE/BTE aids shown in Table 7, ITE aids were given to a higher proportion of males (56%) as compared to females (52%), and BTE aids were given to a higher proportion of females (48%) than males (45%). Similarly, using Naïve Bayesian analysis in Table 7, the likelihood ratios obtained for BTE aids for male and female were 0.92 and 1.08 respectively confirming that more females prefer BTE aids.

For the perception of cosmetics, BTE aids nowadays may be less noticeable, while ITE aids are easier to see from front and side facing. The audiologist informed us that about 70% of patients express a preference for ITE, after seeing a leaflet with the pros and cons of ITE/BTE aids. This preference for ITE aids is significantly greater in males, since perhaps men are more vain than women, especially older ones.

Table 7. Sex with observed frequencies and likelihood ratios of ITE and BTE aids.

Sex	ITE	BTE	Row total	Likelihood ratios	
				ITE	BTE
Male	1496 (56%)	1188 (45%)	2684 (47%)	1.09	0.92
Female	1577 (52%)	1475 (48%)	3052 (53%)	0.93	1.08
Column total	3073 (54%)	2663 (46%)	5736 (100%)		

6. RE-RUN OF EXPERIMENTS

6.1 Naïve Bayesian analysis experiments after removing BC250 Hz and mould

After considering the comments and feedback from the audiologist, the models for the audiology decision support system were constructed again after removing the factors BC250 Hz and mould. The audiologist commented that the mould field should be left blank, as some moulds are peculiar to one type of aid, and thus are wholly predictive of that type of aid (BTE). Then, Naïve Bayesian analysis produced similarity was in the range 63 to 66%, precision was in the ranges 0.63 to 0.67 for ITE and 0.62 to 0.65 for BTE, recall was in the ranges 0.71 to 0.73 for ITE and 0.53 to 0.60 for BTE, and the F-measure was in the range 0.67 to 0.69 for ITE and 0.57 to 0.62 for BTE. These results show that for each cross validation run, both the final model and the success rates were similar.

7. CONCLUSION

In this research the authors have examined and analysed for factors influencing the choice between two common hearing aid types: BTE (worn behind the ear) or ITE (worn in the ear). This choice is typically made by audiologists working in out-patient clinics, on the basis of audiogram results and in consultation with the patient. In many cases, the choice is clear cut, but at other times the audiologists might benefit from a second opinion given by an automatic system with an explanation of how that second opinion was arrived at. Such a decision support system is developed in this research using the technique of Naïve Bayesian analysis. In this research, candidate factors found were combined using their likelihood ratios into a single model to predict the outcome of hearing aid type, for example it was found in Table 7 that females prefer BTE aids and males prefer ITE aids. Also, some of our intermediate results are of interest in themselves, for example, ITE aids are more often prescribed to patients in 60-78 age group and patients more than 78 years of age are given BTE aids as shown in Table 6.

This work has used an unconventional method (data mining) to predict outcomes in audiology for the research question. Data mining of audiology data is a good approach because it finds all possible associations in the data then the most significant associations are filtered. Research in audiology based purely on data mining has not been performed before in the literature, because most previous studies did not have access to such a large amount of data of such richness as was used in this study. The main previous studies in the data mining of audiology records [2] used entirely different approaches, those of rough set theory [2] and decision trees [4]. These two studies were much smaller than the current study, using 500 and 550 records respectively. The DSS were constructed from a large heterogeneous audiology data set. Heterogeneous data is typical of medical records [17], which consists of records typical of a relational database (such as age, gender, or diagnosis), different graphs, free text data, and images. Although this data set does not include images, the techniques developed in this research are thus capable of processing medical data in general. The discovery of relationships in audiology data and DSS will provide supplementary information for audiology experts and hearing aid dispensers. This type of system may also be of interest to manufacturers of hearing technologies in using as a ready means for their telephone customer services staff to check data, discovering data in audiology records will also be good for general awareness about the suitability of hearing aid type.

8. ACKNOWLEDGMENTS

The authors would like to thank the concerned people at a large NHS facility for making the large set of anonymised audiology records available to them. They would also like to thank the principal audiologist of this NHS facility for giving his comments and feedback on their work.

9. COMPETING INTERESTS

The authors report no conflict of interest.

10. STUDY APPROVAL

It is a study involving anonymised audiology data NOT human directly. This data was given to Department of Computing, Engineering & Technology, University of Sunderland, Sunderland, UK, for Research purpose by a large National Health Services facility after making the data anonymised. So, there was no need to obtain any permission as long as it is being used for research purpose.

11. REFERENCES

- [1] Ananiadou, S. 2009. Text mining for biomedicine. In: Prince, V., Roche, M., eds. Information retrieval in biomedicine, natural language processing for knowledge integration. Medical information science reference, IGI Global, 1-9.
- [2] Bakar, A. A., Othman, Z., Ismail, R., Zakari, Z. 2009. Using the rough set theory for mining the level of hearing loss diagnosis knowledge. International conference on electrical engineering and informatics; Selangor, Malaysia, 7-11.
- [3] Shalvi, D, DeClaris, N. 1998. An unsupervised neural network approach to medical data mining techniques. Neural Networks Proceedings, IEEE World Congress on Computational Intelligence, 171-176.
- [4] Thompson, P., Zhang, X., Jiang, W., Ras, Z. W. 2007. From mining tinnitus database to tinnitus decision support system, initial study. IEEE/WIC/ACM International conference on Intelligent Agent Technology, 203-206.
- [5] Deafness Research UK Annual Review (dated: 6th August 2009) <http://www.deafnessresearch.org.uk/Deafness%20Research%20UK%20publishes%20Annual%20Review+5294.twl> (accessed 6 July 2011)
- [6] Bertoli, S., Staehelin, K., Zemp, E., Schindler, C., Bodmer, D., et al. 2009. Survey on hearing aid use and satisfaction in Switzerland and their determinants. *Int J Audiol*,48:183-195.
- [7] NHS: How can I get an NHS hearing aid? (dated: 9th February 2011) <http://www.nhs.uk/chq/Pages/894.aspx?CategoryID=68&SubCategoryID=157> (accessed 8 July 2011)
- [8] Wong, L. L. N., Hickson, L., McPherson, B. 2009. Satisfaction with hearing aids: A consumer research perspective. *Int J Audiol*, 48:405-427.
- [9] Tyler, R., Coelho, C., Tao, P., Ji, H., Gehringer, A., et al. 2008. Identifying tinnitus subgroups with cluster analysis. *Am J Audiol*,17(2):76-184.
- [10] Andersson, G., McKenna, L. 1998. Tinnitus Masking and Depression. *Int J Audiol*, 37:174-182.
- [11] Anwar, M. N., Oakes, M. P., Wermter, S., Heinrich, S. Clustering Audiology Data. In 19th Annual Belgian-Dutch Conference on Machine Learning (BeneLearn 2010); Leuven, Belgium 2010. http://dtai.cs.kuleuven.be/events/Benelearn2010/submissions/benelearn2010_submission_7.pdf (accessed 24 June 2012)
- [12] Anwar, M. N., Oakes, M.P., McGarry, K. 2011. Chi-squared, Yule's Q and Likelihood Ratios in Tabular Audiology Data. In: Ao, S. L., and Gelman, L., eds. Electrical Engineering and Applied Computing. Dordrecht: Springer Netherlands, 465-376.
- [13] Witten, I.H., Frank, E., Hall, M. A. 2011. Data mining, practical machine learning tools and techniques. Third Edition. Morgan Kaufmann Publishers, 124-126,261-273.
- [14] Han, J., Kamber, M., Pei, J. 2012. Data mining: concepts and techniques. Third Edition. Morgan Kaufmann Publishers, 350-351,393 -396.
- [15] Manning, C.D., Raghavan, P., Schütze, H. 2008. Introduction to information retrieval. Cambridge University Press, 142-144.
- [16] Manning, C.D., Schütze, H. 1999. Foundations of statistical natural language processing. The MIT Press, 232-241.
- [17] Cios, K.J., Moore, G. W. 2002. Uniqueness of medical data mining. *Artif Intell Med*, 26:1-24.