JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 36, 205-215 (2020)
DOI: 10.6688/JISE.202003_36(2).0002

Applying Datamining Techniques to Predict Hearing Aid Type for Audiology Patients

MAALIM A. ALJABERY[1,2] AND SEFER KURNAZ[1]
[1]*Faculty of Engineering and Natural Sciences*
Electrical and Computer Engineering Department
Altinbas University
Istanbul, 34676 Turkey
[2]*Faculty of Science, Computer Science Department*
University of Basrah
Basrah, 61004 Iraq
E-mail: maalim.aljabery@ogr.altinbas.edu.tr; sefer.kurnaz@altinbas.edu.tr

Our research is primarily based on dealing with different types of data using Data Mining (DM) techniques. In this research, we devoted ourselves to determining the type of Hearing Aid (HA) needed by patients with hearing impairment.

HA type Diagnosis is a medical application that is a major challenge for researchers. Using DM techniques and Machine Learning (ML) has created a major challenge in the process of predicting the appropriate HA type for Audiology Patients (APs). Thus, this problem is primarily in the domain of classification problems. Our study makes a summary of some technical articles on determining the specific type of HA and introduces a study of using DM techniques to improve the accuracy predict for this purpose.

Furthermore, our research includes the creation of a new Audiology Dataset based on the addition of some important fields on the old audiology database and analyses a new data of APs. These data have been obtained from the field work for nearly eight consecutive years, then extract a new classification based on this analysis.

Relied on our search to reach the highest degree of accuracy in predicting the type of appropriate HA for APs who use it to enhance their hearing, we applied, compared, and analyzed the Neural Network (NN) and Support Vector Machine (SVM), applying Anaconda Navigator version 1.7.0, Orange Canvas version 3.13.0, and Spyder version 3.2.6 applications for Python coding.

Keywords: data mining, hearing aid, audiology patient, neural network, support vector machines

1. INTRODUCTION

Data in general have a very important role in human life. **DM** is the process of knowledge discovery through the analysis of large amounts of data from different perspectives and extract beneficial information [1].

DM in the current framework is possess an enormous spectrum, that has encouraged researchers to comprehend linear and nonlinear correlations [2].

Recently, has been rapidly progressing in the applications of **DM** with regard to audiology data and has expanded significantly to include most cases of hearing diseases [3].

The predictive modelling is one of the importance **DM** objectives which is essential and considered as a frequently applied function [4].

Received September 7, 2019; revised September 11 & 26, 2019; accepted October 11, 2019.
Communicated by Osamah Ibrahim Khalaf.

The Information Technology in healthcare field implicates authorized of creating electronic records for patients from surveys of their visits [5].

Medical data refer to the databases that contain information about health care, like the records of the patients, and with the development technology of information, a lot of medical data are saved electronically [6].

Applying "Data Mining" techniques and its tools to establish a clear visualization of the combination of patients and their hearing problems in the class of audiology dataset and to find the new inferences from utilizing individual data with the aggregation of datasets which will instructs to individual potentials and improve the real-world potentials rather than the laboratory procedures for selecting the **HA** type. The difficulty of these inferences will differ, and in the beginning, the results to be announced should be confirmed from laboratory studies, such as adaptation to the work of the device. As a test of the efficiency algorithms utilizes, it is preferable to have new inferences to be developed related to the field of research, and as researchers in this field, we will explore how **DM** will be able to develop the required tools and solutions to learn profound areas of human biology and the importance of **DM** that will be provided when developing the algorithms and procedures for the healthcare system. So, we will concentrate on previous studies related to select an appropriate **HA** and how we can utilize the audiology data to support audiologist an accurately evaluations for their patients. Furthermore, we will study some of the researches done to consider the advantages and disadvantages of these studies which conducted using **DM** techniques for predicting the type of **HA** for **Ap**s. Most important of these studies are:

- Data mining of audiology patient records: factors influencing the choice of hearing aid type [7].
- Data Mining Audiology Records with the Chi-squared Test and Self-Organising Maps [8].
- Decision Support System for the Selection of an ITE or a BTE Hearing Aid [9].
- Predict the type of hearing aid of audiology patients using data mining techniques [4].
- Hearing aid classification based on audiology data [10].
- Generating and using "big data" to identify hearing aid patterns of usage in order to optimise and personalise fitting [11].
- Proposed audit of provision of hearing aids for mild/moderate hearing losses [12].

The imperatives and challenges identified with security affectability and the vast volume of restorative information assume imperative part in the determination of the specific information mining strategy. Also, moral and legitimate parts of restorative information are likewise vital viewpoints. Restorative information can have an extraordinary status in view of its pertinence to all individuals [13].

This research submits a portion of **DM** and its utilization in audiology fields and it characterizes the **DM** technology as well as its applications which would be helpful in the fields of power systems. The goal is to build a network connecting the developers, audiologists, and agents from manufacturers of **HA**s and clinical diagnostic tools to know how they can utilize patient data to build a comprehensive visualization for users.

Finally, we used Orange Canvas for modelling, design our model utilizing Data Mining techniques, and coding via Python.

2. METHODS AND ANALYSIS

We created a new dataset for the preservation of specific data related to the diagnosis of patients with hearing problems.

The datasets should be chosen accurately and pre-processed if it needs to, also it should be authoritative, dependable, and applicable, since some of data types are not suitable for classification methods or decision tree learning to be definite [4].

Our dataset consists of 71 fields and one extra field for class. All fields of this new dataset are categorical with little ratio of missing values. Regarding to the actual data, we obtained the data of 17,300 patients by working in the Hearing and Speech Evaluation Unit of Basrah General Hospital in Iraq and from Al Nour clinic for Hearing and Speech Audiometry in Basrah for 8 years of continuous work in these places. After cleaning data, we were obtained 210 cases ready for exploration using the techniques of **DM** to predict the type of **HA** needed by **APs**.

Table 1 illustrates the summary of our new dataset for APs.

Table 1. The summary of the new audiology dataset.

Source	Dataset	No. of Instances	No. of Classes	No. of Attributes	Missing value %	No. of Nominal
Owned by the researcher	Specific Audiology	210	3	71	2.8	71

2.1 DM Techniques

In our study, we will address two of **DM** algorithms for prediction and comparison. These algorithms are **NN** and **SVMs** as listed below:

- **Neural Network (NN):** The **NN** is a based approach to mining the rules of classification from the database which has worked with. The initial stage has trained the network to achieve the required rate of accuracy. Then removed the network connections which redundant via algorithm pruning of the network. After that, the hidden nodes activation values for this network will be analyzed and finally, generate the rules of classification by using the analysis outcomes [14].
- **Support Vector Machine (SVM):** The first introduced of **SVMs** was by Vapnik. It depends on the binary classification, so it isolates the vectors training set into two various classes (a_1, b_1), (a_2, b_2),. ., (a_m, b_m), where $a_i \in R^d$ which indicate to the vectors of feature space of d-dimensional, and $b_i \in \{-1, +1\}$ is a label of the class [15].

In order to achieve the desired goal, we used the Anaconda application. Then, we designed the main model using Orange Canvas, and we extracted the required results using the Spyder programming in Python (they are already implicit in Anaconda), as listed below:

- **Anaconda Navigator:** A lot of software packages and versions of **DM** are supplied applications for decision tree algorithms. We are implemented our model using Anaconda Navigator version 1.7.0, since it contains other applications that we need to imp-

lement our work, such as Orange Canvas (for modeling) and Spider (for Python coding).

- ❖ **Orange Canvas (version 3.13.0):** Orange is a component-based **DM** framework, data visualization and analysis for novice and expert, and interactive workflows with a long toolbox. It is a multilayer architecture of a general-purpose **DM** tools and machine learning technique. It's utilizing for data analysis via Python and visual programming [16].
- ❖ **Spyder (version 3.2.6):** Spyder is a scientific Python Development Environment. It is a powerful Python Integrated Development Environment **(IDE)** with advanced editing, interactive testing, debugging, and introspection features.

Finally, we obtained accurate results for the data training stage. The method was then circulated to the data testing stage for **APs**.

2.2 The Model Designs

The design of our model applied using Orange Canvas modeler (Fig. 1).

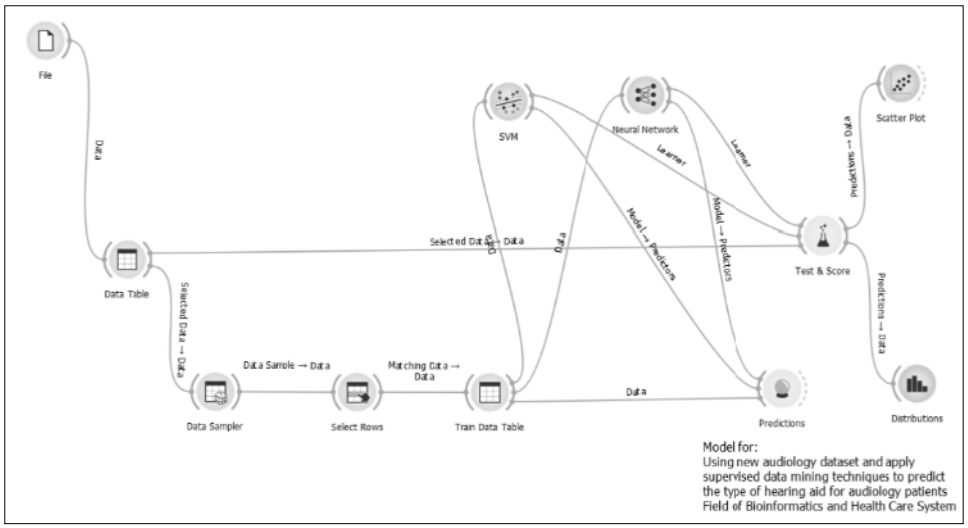


Fig. 1. The design of a graphical interface for our model using Orange Canvas.

2.3 Evaluation of Model Performance

Table 2 shows the Test and Score evaluation results of Support Vector Machine (**SVM**) and Neural Network (**NN**) algorithms on training data using Orange Canvas for Behind The Ear (**BTE)** Hearing Aid (**HA)** target class as a class prediction result.

Table 2. Test and score of algorithms on train data for behind the ear (BTE) target class.

Method	AUC	CA	F1	Precision	Recall
SVM	1.000	0.996	0.992	0.985	1.000
Neural Network	1.000	1.000	1.000	1.000	1.000

Table 3 shows the Test and Score evaluation results of Support Vector Machine (**SVM**) and Neural Network (**NN**) algorithms on training data using Orange Canvas for In The Ear (**ITE)** Hearing Aid (**HA)** target class as a class prediction result.

Table 3. Test and score of algorithms on train data for in the ear (ITE) target class.

Method	AUC	CA	F1	Precision	Recall
SVM	1.000	1.000	1.000	1.000	1.000
Neural Network	1.000	1.000	1.000	1.000	1.000

The mean rank of error rate and mean error rate are the way to measured classification precision [17].

Table 4 shows the predictions details for the two classes **BTE** and **ITE HA**s and the comparisons of results for both algorithms **NN** and **SVM**.

Table 4. The predictions details for behind the ear and in the ear hearing aids for neural network and support vector machine algorithms.

Predictions Information	Index	NN Predictions			SVM Predictions		
		BTE Pred.	ITE Pred.	NN Result	BTE Pred.	ITE Pred.	SVM Result
Data: 12 instances Predictors: 2 Task: Classification Predicted Probabilities for: BTE and ITE Classes Condition: If he_class is not normal_heari Data: In: ~20 rows, 72 variables Out: ~12 rows, 72 variables Data Sampler: 210 instances in input dataset Outputting 20 instances	1	0.05	0.92	ITE	0.15	0.80	ITE
	2	1.00	0.00	BTE	0.99	0.00	BTE
	3	0.01	0.98	ITE	0.03	0.96	ITE
	4	0.00	0.99	ITE	0.01	0.98	ITE
	5	0.99	0.01	BTE	0.97	0.02	BTE
	6	0.05	0.92	ITE	0.15	0.80	ITE
	7	0.99	0.01	BTE	0.99	0.00	BTE
	8	0.00	1.00	ITE	0.01	0.97	ITE
	9	0.98	0.01	BTE	0.95	0.01	BTE
	10	0.99	0.00	BTE	0.97	0.01	BTE
	11	0.99	0.00	BTE	0.98	0.01	BTE
	12	0.00	0.99	ITE	0.01	0.98	ITE

2.4 Discussion and Analysis

Neural Networks (**NN**) and Support Vector Machines (**SVM**) are most popular techniques for classification and supervised Machine Learning (**ML**). It is often not clear which technique is much better for a specific project. So, there are some aspects with an **NN** such as Multilayer Perceptron and specifically which might prefer it as likeable to use rather than **SVM**. On the other hand, there are some aspects make **SVM**s are more superior than **NN**s since they prevent two significant weaknesses that **NN**s have:

- In many cases there is a converge in **NN**s on local lowest level comparing with global lowest level, mostly that mean they are basically, missing the trees forest in some cases.

- Mostly, there is an overfit of the **NN**s if the time of training is too long, that is mean for any pattern will give, the start of **NN** might take into consideration the noise and dealing with it as a pattern part.

These two hard troubles are not happening with **SVM**s. Regardless, it's not clearly that **SVM**s are apparent refer to be the completely alternate for **NN**s. Thus, what the particular feature(s) that make the **NN**s applicable for confirmed situations rather than **SVM**?

One of the main characteristics of **NN**s that make it much better than **SVM**s is the **NN**s may have an outputs more than one, whereas **SVM**s have just one output.

Moreover, the **NN** will might make further sense as it is one entire, whilst the **SVM**s are isolated and separated systems and this is so benefit when the outcomes are correlated.

The capability of understanding is usually reducing with enlarging of tree growing and complexity. So, when two trees are subjected to the same type of testing then they have the same accuracy of prediction, the smaller leaves tree it should be preferred, when neglecting the training time [18].

3. OVERVIEW

The studies show that people suffering from hearing loss usually complain about having troubles with speech comprehension, mostly when there's a background noise interfering with the sound of the person speaking. The developments and improvements of the technology in **HA** shows that the difficulty of not being able to distinguish speech by the individuals suffering from the hearing loss has been minimized using the outcome which measured by a self-assessment of the **HA**.

Implementation the techniques of **DM** has been activated via the development of data in the areas of electrical systems, these techniques are the multi-task needs that correspond the computer and electrical engineers to a mediator between **DM** and electrical systems [2].

Electrical systems are the significant artificial power systems, which secure, reliable, provident and perform a highly significant portion in guaranteeing the development of the socioeconomic. Thus, some modern **DM** techniques and new technologies have been inserted in this area [19].

With progressively environmental qualifications the recent electrical systems are examined big transformations in numerous ways. The data quantity in the electrical systems is rapidly growing due the huge use of the database of electrical engineers' systems for diverse operations. From the generation of power plants, the energy of electrical transferred for users [11].

The impact of data has an essential part in the Information Technology (**IT**) area. The prime task of the **DM** techniques is to converge data and information from diverse sources and then save, analyses, and preserve it to gain the required information.

The significant reasons for applying **DM** in the scope of the electrical system are to discover the important patterns in data that contain diverse proceeding of the operating systems [2].

Lately, the desires to use information wasted are continuous to make it influential and effective enough to the big amount of data generated from the operations of the system. Thus, **DM** techniques deal with a significant role to work in the control and operating of the electrical system. These **DM** techniques are utilized to discover out unknown relationships of data which make predictions are available [7].

The preparation and operation of electrical systems supply a huge quantity of data that is hard to excerpt and extract helpful information of this huge database which is always utilized.

DM is a process of excerpting and extracting the knowledge that unknown previously from any dataset. The techniques of **DM** assist the planners and operators of electrical systems to gain facilitate planning of operating systems and are helpful for excerpting beneficial information that banks of data are presented [8].

Many researches are conducted for the classification of data loaded. It enables clustering via supervised and unsupervised methods. Unsupervised techniques like Self Organizing Map, k-means, Euclidean range of classify loads, hieratical, and density clustering [20]. Whereas Supervised learning is a technique of **ML** for generating a function of data for training. This training data contain a set of the required output and the input data [21].

Supervised techniques contain Support Vector Machines, Neural Networks, Naive Bayes, Logistic Regression, and Decision Tree [20].

Supervised classification (which related to our research topic) has prevailed in current researches and recently, has been rapidly progressing in the applications of **DM** with regard to audiology data and has expanded significantly to include most cases of hearing diseases [3].

4. EDUCATIONAL MERITS

Our approach is the application of data mining which is present the new technology that evolve and generate the new application axis which is called **Educational Data Mining** that deals with tasks like:

- Prediction includes density prediction, regression, and classification.
- Clustering.
- Mining of the relationships, includes connections, relations, consecutive mining, and vocalic DM.
- Data distillation for judgment of a human.
- Models discovery.

Furthermore, by applying **Educational Data Mining**, the educators enable to handle both classical and cleared educational subjects and gain an advantage from the valid decision making. Moreover, **Educational Data Mining** goals are developing the comprehension and supervision of performance domain of administrators, teachers', students', instructional correlations, and behaviors [20].

DM has extraordinarily attracted numerous businesses related with the predictive and descriptive susceptibility it pledges, such as **Education** of its wide domains for the

hierarchies of it's organizational. Currently, **Education** involves the knowledge and inherited information of communication and transmit, also the **Education** has become a global and comprehensive business with immoderate requirements in information dealing, analysis, and treating [19].

A new database and new audiology dataset based on the good results we obtained with regard to the high predictive rate of choice the type of **HA** for **APs**, the main educational merits of this paper are that it can become an important reference that can be adopted by the researchers at all levels, from undergraduate studying, postgraduate, and expert researchers in this area.

In addition, this work will provide a great service for specialists in the field of data mining and also will be considered for audiologist and specialists doctors in the field of hearing and balance, and certainly will be an important supports for patients who need to guide them to choose the type of **HA** that suits their cases.

Furthermore, this research paper is to learn the impact that **DM** has already had on the medicine. This field produces data considered to be extremely huge, and that should have a way to be analyzed efficiently, effectively, and reliability such as most institutions that create a lot amount of data.

The potentials of making data understandable while centralizing and making them simply readable and reachable by medical audiologists and also can be helpful and beneficial. The study given here explains the effect by showing how **DM** is able to have the specialist doctors ability when prescribing the **HA**, and it also promises the continuous influence of **DM** over the medicine progression and the treatment of patients which done by the audiologists.

5. CONCLUSION

In the scope of medicine, a lot of incoming data are continuously generated and usually, the produce of data is more higher comparing with the produce of the knowledge amount. Thus, high augmentation in data production needs rapid transportation to knowledge. To reach this goal, the set's items of data should examine by clustering to determine the similarity or dissimilarity of this data. It perceives the areas of rising density sample and displays the clusters centers. The technique of clustering contain the approaches of artificial **NN** [22].

The variables such as tympanogram, audiogram, gender, age, **HA** type, etc. are regarding the selection of **HA** type by utilizing the **DM** techniques [23].

Thus, allow the audiologist to construct a proper decision about whether utilizing the original diagnoses or reappraise, while it includes in the classification the attributes details that have the most weighed in the prediction. Besides, there are some considerations may the audiologist take it into account since they persist to refine further data, certainly, that will add more confidence for clinics when giving the correct choice about choosing the type of **HA** for their patients.

To achieve the desired objectives in achieving the highest prediction rate, the technologies provided by **DM** are utilized to rely on the best and most accurate results.

The use of **HA** offers some advantages to a person with hearing impairment. First of all, they will hear more better. Besides, **HA** do not fully recover hearing, but it en-

hances hearing greatly. With using **HA**, the speech of others becomes easy to hear and sounds that **APs** have not recognized for a long time like singing birds, ringing bells, knock doors, wind whistling and running water, it usually become familiar to them.

Furthermore, utilizing the **HA**s for people who suffer from hearing loss are very helpful for the clarity hearing and speech understanding. So, to guarantee the full performance of the **HA** device, it was needful to utilize the experiences of the audiology specialists who conduct the hearing evaluation and adjust the settings of **HA** system.

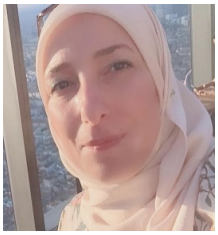
In our research, we will compare some **DM** techniques to mine the data of **AP**s then make the decisions concerning patients, besides we can examine how these techniques can be helpful in the field of the healthcare system.

Finally, our research includes the mechanism of employing **DM** techniques to obtain the best prediction rate by comparing the results obtained and thus determining the best technique for selecting the type of **HA** for people with hearing impairment.

REFERENCES

1. M. Mahamune, S. Ingle, P. Deo, and S. Chowhan, "Healthcare knowledge management using data mining techniques," *Advances in Computational Research*, Vol. 7, 2015, pp. 274-278.
2. A. K. Mishra and A. K. Saxena, "Data mining technology and its applications to power systems," *International Journal of Computer Applications*, Vol. 131, 2015, pp. 1-6.
3. J. Mellor, M. A. Stone, and J. Keane, "Application of data mining to a large hearing-aid manufacturer's dataset to identify possible benefits for clinicians, manufacturers, and users," *Trends Hear*, Vol. 22, 2018, pp. 1-15.
4. S. Kurnaz and M. Aljabery, "Predict the type of hearing aid for audiology patients using data mining techniques," in *Proceedings of the 4th International Conference on Engineering & MIS*, 2018, No. 57.
5. E. M. Beulah, S. N. S. Rajini, and N. Rajkumar, "Application of data mining in healthcare: A survey," *Asian Journal of Microbiology*, *Biotechnology and Environmental Sciences*, Vol. 18, 2016, pp. 999-1001.
6. S. C. Pandey, "Data mining techniques for medical data: A review," in *Proceedings of International Conference on Signal Processing*, *Communication*, *Power and Embedded System*, 2017, pp. 972-982.
7. M. N. Anwar and M. P. Oakes, "Data mining of audiology patient records: Factors influencing the choice of hearing aid type," *BMC Medical Informatics and Decision Making*, Vol. 12, 2012, pp. 1-8.
8. S. Cox, M. Oakes, S. Wermter, and M. Hawthorne, "AudioMine: Medical data mining in heterogeneous audiology records," *International Journal of Computer and Information Engineering*, Vol. 1, 2007, pp. 141-144.
9. M. N. Anwar and M. P. Oakes, "Decision support system for the selection of an ITE or a BTE hearing aid," *International Journal of Computer Applications*, Vol. 76, 2013, pp. 37-42.
10. C. Panchev, M. N. Anwar, and M. Oakes, "Hearing aid classification based on audiology data," in *Proceedings of International Conference on Artificial Neural Net-*

works and Machine Learning, 2013, pp. 375-380.
11. J. C. Mellor, M. A. Stone, and J. Keane, "Application of data mining to 'Big Data' acquired in audiology: Principles and potential," *Trends Hear*, Vol. 22, 2018, pp. 1-10.
12. H. C. Koh and G. Tan, "Data mining applications in healthcare," *Journal of Healthcare Information Management*, Vol. 19, 2005, pp. 64-72.
13. D. A. Abdulameer, "Medical data mining: Health care knowledge discovery framework based on clinical big data analysis," *International Journal of Scientific and Research Publications*, Vol. 5, 2015, pp. 1-6.
14. S. M. Kamruzzaman and A. M. J. Sarkar, "A new data mining scheme using artificial neural networks," *Sensors*, Vol. 11, 2011, pp. 4622-4647.
15. M. W. Huang, C. W. Chen, W. C. Lin, S. W. Ke, and C. F. Tsai, "SVM and SVM ensembles in breast cancer prediction," *PLoS One*, Vol. 12, 2017, pp. 1-14.
16. J. Demsar and B. Zupan, "Orange: Data mining fruitful and fun − A historical perspective," *Informatica*, Vol. 37, 2013, pp. 55-60.
17. M. Bekkar, H. K. Djemaa, and T. A. Alitouche, "Evaluation measures for models assessment over imbalanced data sets," *Journal of Information Engineering and Applications*, Vol. 3, 2013, pp. 27-39.
18. J. K. Martin, "An exact probability metric for decision tree splitting and stopping," *Machine Learning*, Vol. 28, 1997, pp. 257-291.
19. Y. Zhang, J. Ma, J. Zhang, and Z. Wang, "Applications of data mining theory in electrical engineering," *Engineering*, Vol. 1, 2009, pp. 211-215.
20. C. E. Moucary, "Data mining for engineering schools," *International Journal of Advanced Computer Science and Applications*, Vol. 2, 2011, pp. 1-9.
21. B. Twala, M. Cartwright, and M. Shepperd, "Applying rule induction in software prediction," *Advances in Machine Learning Applications in Software Engineering*, 2006, pp. 265-286.
22. N. Anusha, Rajashree, K. S. "Survey on medical data by using data mining techniques," *International Journal of Science, Engineering and Technology Research*, Vol. 7, 2018, pp. 11-17.
23. M. N. Anwar, "Mining and analysis of audiology data to find significant factors associated with tinnitus masker," *Springerplus*, Vol. 2, 2013, pp. 1-10.



Maalim A. Aljabery received the B.S. degree in Computer Engineering and M.S. degree in Computer Science Department from University of Basrah, Iraq. She is currently pursuing the Ph.D. degree in the Department of Electronic and Computer Engineering, Altinbas University, Turkey. Her research interests include data mining and knowledge discovery, hearing aids devices, and audiology (field of medicine related with patients who suffer from hearing problems in general).

Sefer Kurnaz was born in April 19, 1956 in Giresun, Turkey. He has a retired colonel from Turkish Air force Command. In the past, he was in charging of as Director of Aeronautics & Space Technologies Institute (ASTIN) in Turkish Air force Academy, Istanbul. Sefer Kurnaz has mainly specialized in space technology application. He has graduated Ph.D. at the İstanbul University and specialized in computer science. He was worked at 6 Allied Air Tactical Force (6ATAF) in NATO, Izmir as a system support section chief. He is a founder and before retiring was a general chair Resent Advances in Space Technologies (RAST) International Conference, Turkey. He is an author of two books in the famous publishers in Turkey and United States of America. Under his leadership have published over than 20 scientific papers.