Predict the type of hearing aid for audiology patients using data mining techniques

Asst. Prof. Dr. Sefer Kurnaz
Faculty of Engineering and Natural Sciences
Altinbas University
Istanbul, Turkey
sefer.kurnaz@altinbas.edu.tr

Maalim A. H. Aljabery
Electrical and Computer Engineering
Altinbas University
Istanbul, Turkey
maalim.aljabery@ogr.altinbas.edu.tr

***Abstract*— Our research transacts with a great various audiology data from National Health System (NHS) facility, including audiograms, structured data such as age, gender, and diagnosis, and a text of specific information about each patient, i.e., clinical reports. This research examines factors related to audiology patients depends on various data by using the mining and analysis of this data. This paper looks for factors affecting the choice between two prevalent hearing aid kinds: BTE (Behind The Ear) or ITE (In The Ear). This choice often done by audiology technicians working in specific clinics for this purpose, based on audiograms results and patient consultation. In many situations, there is an obvious choice, but sometimes the technicians need for the second opinion via an automatic system includes clarification of how to obtain that second opinion. The research deals with diversified specifics and more significant factors for choosing of confirmed hearing aid related to those specifics. We depend on the earlier study data (Bareiss, E. Ray, & Porter, Bruce (1987)). Protos: An Exemplar-Based Learning Apprentice. In the Proceedings of the 4th International Workshop on Machine Learning, 12-23, Irvine, California, which illustrates the database analysis for 180,000 records, for 23,000 patients, by the hearing aid clinic at James Cook University Hospital in Middlesbrough, UK. This data mined to find which factors contribute to the deduction to fit a BTE hearing aid as opposed to an ITE hearing aid. Here we conduct some enhancements on this database and analyze the data depends on medical information to create a new class then we use some intelligent Data Mining (DM) techniques to guess the most correct illness that could be associated with patient's information. Based on the result (according to the patients' diagnosis details), we can obtain right predictions of which type of Hearing Aid (HA) they should use.**

Keywords*— audiology, National Health System, audiograms, BTE, ITE, Machine Learning, Data Mining, Hearing Aid.

I. INTRODUCTION

Data mining is one of computer science areas which deals together with:

a) Artificial intelligence conception. b) Data structures. c) Statistics. d) Databases [1]. Data mining is the concealed predictive information that extract from enormous databases. There are diverse kinds of data mining objectives. However, there are two of them are most commonly used:

1) predictive modelling.

2) descriptive modelling (is out of this search).

Predictive modelling is fundamental, since through this function the predictor can create predictions about the future through learning from the previous data. Within the concept of data mining, this should be considered as a frequently applied function. The tools of data mining are predicting future tendencies and demeanours, authorizing users to create proactive, knowledge-driven resolutions [2]. Classification is a data mining mechanism used to predict a set of membership for data samples [3]. It can treat enormous diversity of data than retrogression and increasing in popularity [3]. Data mining specifically clustering techniques lets inspecting and comparison the similarity and dissimilarity between datasets. Clustering techniques contain artificial neural network approaches that qualify to collect a set of data in groups to be more like each other comparing with those in other groups [4].

With the employ of data mining techniques, it is possible to extract good and beneficial knowledge, information, and regularities [5]. It is a highly needed area since many institutions and companies can assist and gain benefits from it. Healthcare organizations nowadays are eligible for generating and combining a great amount of data. This rise in the size of data requires an automatic method for these data to be extracted when have to.

Certainly, that it is a magnificent idea to use data mining instruments in the healthcare systems to enhance decisions made by the specialists in this area. It will be large supporting to the healthcare system to learn what the origin of troubles is [1]. Information technologies in healthcare include authorized the creation of electronic patient records that acquired from surveillance of the patient visits. This information contains patient demographics, therapy progress archives, specifics of sponsorship, prescribed drugs, examination results, previous medical history, lab outcomes, etc. [5]. The most significant challenge in this supporting is the case of dealing with huge data and utilization.

Data mining algorithms can be applied to learn beneficial types in the patients' statistical information to extract the correlation among some disease and potential reasons for them. Therefore, healthcare organizations are discovered values and planned implementations to mining the data of patients, and community data, in specific [1].

The main aim of the research is concise with the significance of specifying patient needs that being a challenge with the forward features of the new hearing aid type. People who suffer from hearing problems requests and anticipations rise along the commercial promotion of specific hearing aid features like the microphones sides control, reducing the noise environment, and telecoils adaptive.

In general, hearing aids are most useful in enhancing the hearing and speech understanding of people who suffer from hearing impairment. Indeed, to ensure perfect performance of the type of hearing aid instrument, it was necessary to include the skills of a specialist who performed the hearing assessment and modified the settings of the hearing aid device. This research will study some issues where used the data mining to mine patient data to create decisions regarding patients and inspect how data mining can be advantageous in the backgrounds of the healthcare system.

978-1-5386-5169-8/18/$31.00 ©2018 IEEE

1. The Purpose of Current Approach

Our purpose is to construct a network among developers, specialist doctors, and agents from hearing aids industrialists and clinical diagnostic tools to distinguish how to use patients' data to construct an accomplish visualize of the aid user. This visualizes should consider the kind of hearing loss, the susceptibility of the hearing aid user, the conditions in which the aid should be chosen from the user, and the behavior of the hearing aid user when using this device. With an extra precise profile of the hearing aid user, it will be practicable to supply an extra systematic but individualize fitting by considering appropriate measures of:

- Pattern and area of hearing loss.
- The Cognitive function of the hearing aid user.
- Experience of the hearing aid user with wearing the hearing aid device.

With a view to producing a combination image of the users and their hearing loss in the class of a data set. Applying "Data Mining" instruments from the domain of computer science, the purpose is to produce new inferences from an individual and combinations of datasets which will lead to:

- individualized fittings.
- enhanced characterization of the real-world instead of laboratory execution of hearing aids.

These conclusions will differ in their difficulty and at first should confirm results that have been announcing from laboratory studies, like acclimatization to the working of the device. As an examiner of the adequacy of the algorithms used, there should as well be new inferences created that are behind current thinking. With a view to know the importance that data mining is own and will have when the procedures and algorithms have been improved and optimized for healthcare systems, we will explore some examples of what data mining has found and what they are expected to achieve. First, we will focus on a study was done about detect suitable hearing aids and then how we capable to use that data to assist specialist doctors to accurately treat their patients who suffer from hearing impairment. Second, we will look at research done to study and note the advantages and possibilities that offers. Last, we will see how data mining will be useful as researchers evolve the required instruments and solutions to learn the deepest fields of human biology.

2. The History of using Audiology Dataset with field of Hearing Aids

Below we review some of the previous studies in the field of using Data Mining techniques and Audiology dataset to predict the type of Hearing Aid:

i. Data mining of audiology patient records: factors influencing the choice of hearing aid type [6].

ii. Data Mining Audiology Records with the Chi-squared Test and Self-Organising Maps [7].

iii. Decision Support System for the Selection of an ITE or a BTE Hearing Aid [8].

iv. Hearing aid classification based on audiology data [9].

v. Generating and using "big data" to identify hearing aid patterns of usage in order to optimize and personalize fitting [10].

vi. Proposed audit of the provision of hearing aids for mild/moderate hearing losses [11].

3. Methods and Models

3.1. Data Set

Choosing datasets is extremely important for the experiment because the data that is selected necessarily to be reliable and applicable. The means of applicable data is the data which needs to be in the rectify format and prepared for classification studies. Furthermore, not all kinds of data are appropriate for decision tree learning or classification methods to be definite. Thus, the datasets should be chosen accurately, also it should be pre-processed if it is not previously.

3.1.1. Audiology Dataset

The original Audiology dataset version is given to the UCI ('University of California, Irvine') https://archive.ics.uci.edu/ml/datasets/Audiology+%28Original%29, (this web page also publishes the papers that cite this data set) depot in 1992 by Ross Quinlan. However, the first granter was Professor Jergen from the Medicine Baylor College [12].

The purpose is to dissociate the varied type of audiology unorganized causes and to group them into one of 24 sets or classes to recognize the cases based on audiological disorder causes.

3.1.2. Datasets summary

Table I shows the details of datasets summary. The original Audiology dataset attributes and Class Distribution are published in the web site below:

https://archive.ics.uci.edu/ml/machine-learning-databases/audiology/audiology.standardized.names

TABLE I. DATASETS SUMMARY.

Datasets Summary	Original Dataset	New Dataset
Dataset	Audiology	Audiology for HA
No. of cases	226	226
No. of classes	24	3
No. of attributes	69	71
Continuous	0	0
Categorical Attributes	69	71
Missing values	Yes	Yes
Uniform distribution	No	No

[a.] Summary of original dataset and new dataset

3.2. Algorithms and Evaluation Methods

After studying previous research, we have made some additions to the original database by creating a new class relying on the data analysis based on medical information in this area. Then we tested a set of algorithms using the new Audiology dataset. We got an extremely good result and obtained a set of four algorithms chosen to be tested on various evaluation methods. Finally, we gained a perfect percentage of predictions using Orange Canvas for data mining and modeling.

The algorithms used in our research are:

- CN2 rule induction:

  is a classification technique prepared for the efficacious creation of simple understandable rules of form "if condition then predict class", even in noisy areas [13].

- AdaBoost classifier:

  is a meta-estimator starts by applying a classifier on the main dataset and then apply extra samples of the classifier on that dataset, however, where the weights of wrong classified cases are modified such that subsequent classifiers concentrate more on complicated cases [14].

- Random forests classification:

  are a compound of tree predictors where each tree based on the results of a random vector sampled autonomously and with the same classification for all forest trees [15].

- Logistic Regression:

  is a machine learning algorithm used for binary classification [16].

3.2.1. *Data Description and Pre-Processing*

A numerous of data mining software packages and versions provide applications of many decision tree algorithms. We conduct experiments under the framework of Orange Canvas for data mining and modeling.

3.2.2. *Orange Canvas*

Orange Canvas is a machine learning and data mining combination for data analysis via Python and visual programming. The low-level executions at the hierarchy bottom, such as data filtering, probability evaluation, and attribute classification, are congregate into higher-level algorithms, like learning of classification tree [16].

This makes developers append new implementation functions at any level easily and merge it with the existing code.

4. *Results and Comparison*

4.1. *Models Design*

The design summary of the selected models is illustrating in Fig. 1.

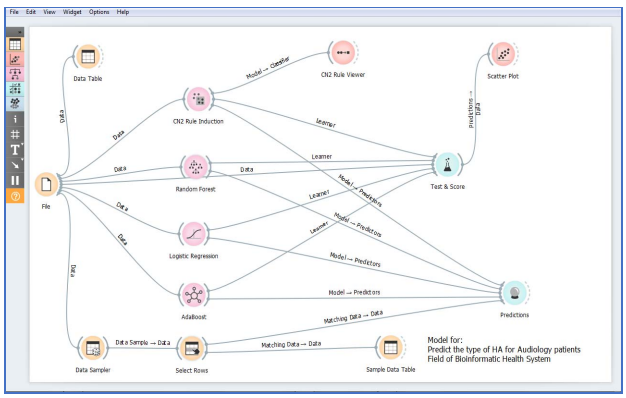

Fig. 1. Screenshoot for the new model using Orange Canvas.

4.2. *Evaluation of Model Performance*

Fig. 2 shows the algorithms "Test and Scores" after evaluation results for "average over classes" selection for 3 classes of the new Audiology dataset using Orange Canvas.

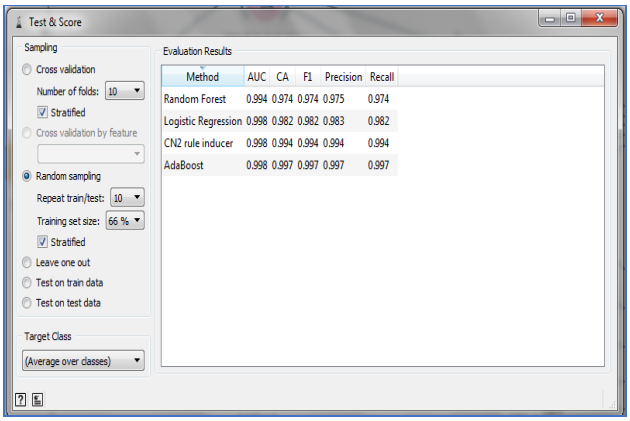

Fig. 2. Screenshoot for the Evaluation results of the Algorithms.

Depends on Target Class:

AUC (Area Under the Receiver Operating Characteristic Curve): Stored predictions and actual data in model testing.

CA: Classification Accuracy score.

F1: Stored predictions and actual data in model testing.

Precision: Is the ratio of tp / (tp + fp) where tp is the number of true positives and fp the number of false positives. The best value is 1 and the worst value is 0.

Recall: The ratio of tp / (tp + fn) where tp is the number of true positives and fn is the number of false negatives. The best value is 1 and the worst value is 0 [17].

While:

$$\text{prediction} = \text{IF (output} < 0.5\text{) Then 0 Else 1}$$

$$\text{Accuracy} = (\text{correct predictions} / \text{number of predictions made}) * 100$$

$$\text{Max Accuracy} = (10/10) * 100 = 100\% \text{ [15].}$$

Classification precision measured via mean error rate and mean rank of error rate. Understandability typically decreases with rising in tree expansion and complexity. If two trees apply the same type of tests and have the same precision of the prediction, the one with smaller leaves is commonly preferred. A third criterion that has been predominantly neglected is the algorithms' relative training time [18].

Fig. 3 shows the "Predictions" done on sample of 25 records selected from the new dataset for all 4 algorithms after modelling and for example, Fig. 4 shows the Scatter Plot for "tymp" (tympanogram) field of a new dataset as X-axis and its attributes as Y-axis. Thus, there are 70 plots for the other fields results for this sample of dataset.

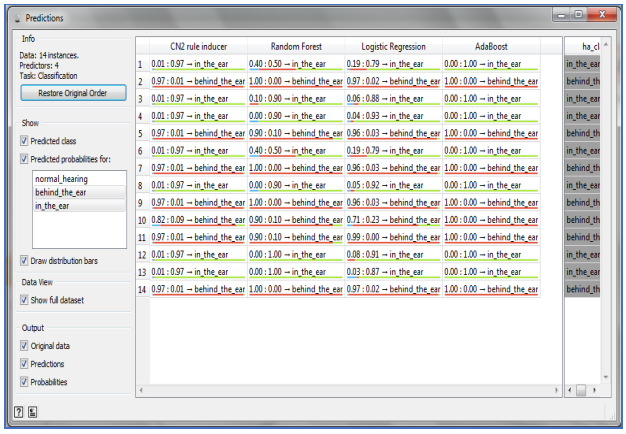

Fig. 3. Screenshoot for the predictions for all 4 algorithms.

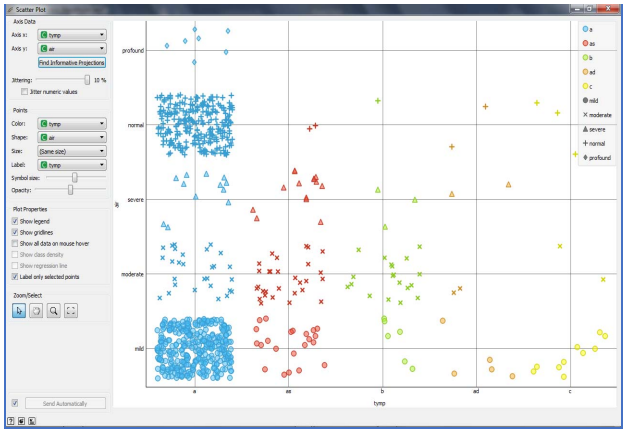

Fig. 4. Screenshoot for the Scatter Plot result of the field (tymp).

4.3. *CN2 conditions*

Before creating new columns on an original Audiology dataset there were 25 conditions in CN2 rule induction as shown in Fig. 5, while after creating a new column for age less than 18 years old and a new class column, we reduced the number of conditions of the CN2 algorithm to just 4 in case of predicting the type of HA, as shown in Fig. 6.

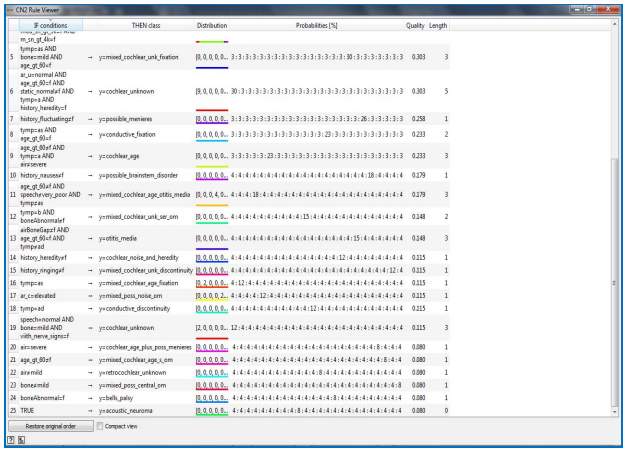

Fig. 5. Screenshoot for the conditions of CN2 rule induction for original dataset.

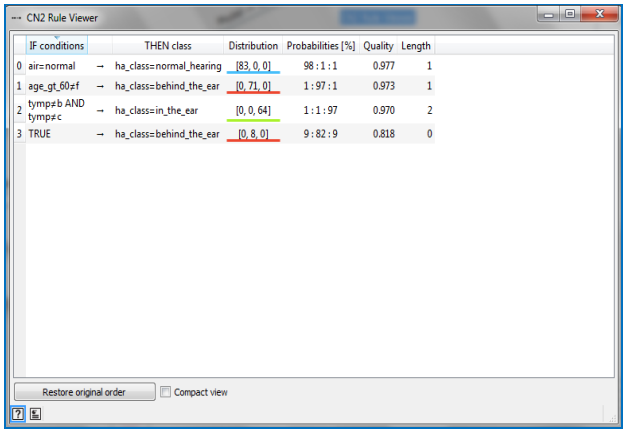

Fig. 6. Screenshoot for the conditions of CN2 rule induction for the new dataset.

5. *Characteristics*

Our research produces a technique that will construct the basis for decision support system of an audiology field, where the hidden records of the patients would be given to the system and would obtain the prediction of which hearing aid the patient should be used with an ITE aid against to a BTE hearing aid.

The advantages of this technique for the collection of evidence is that it is simple to know which variables gave the final decision. Then we should validate these results via obtaining the opinions of specialist persons like a professional audiologist, and by using this technique which builds a model have an interaction between variables [6].

The main advantage of these algorithms which used in the research is they make a clarification ease to be incorporated into the tool of any decision support.

6. *Future development*

By obtaining extra data about the user's hearing impairment, cognitive capacities, acoustic surroundings, and their dynamic conduct within the surrounding, the aid appropriate software, as well as aid software working, can construct up a larger image of both the capacities and incapacities of the user. Future development could drive to the aid dynamically diverse its construct depending on these scales. The operation of explicating how these datasets describe the requirements of the user is critical to enhancing the fixtures of the hearing prosthesis. The employ of data mining techniques is to create a performance of the data and found value by creating further inferences for otherwise invisible relationships inside the data. The purposed effect is that a user is agreeable with using a piece of technology, even so, is unknowing of the complexity that supports it.

For next step in the future work, we will analyze the results with a completely new database and using more than four models for determining the type of hearing aid for hearing impaired patients.

7. *Discussion & Conclusions*

Data mining has magnificent importance in the field of medicine, that represents a comprehensive procedure which required a deep understanding of requirements of the healthcare organizations.

A decision support system was created to predict the kind of hearing aid which should be used, with an illustration facility which illustrates how that decision was reached. This system should obvious beneficial in supporting a "second opinion" for audiologists [6]. Knowledge obtained with the employ of techniques of data mining can be employed to produce effective decisions that will develop the success of healthcare organizations and patients' health. Data mining needs convenient technology and analytical techniques, besides, systems for tracking and reporting which capable measuring of results [4]. The links between audiogram type and hearing aid type were approved by the preceding work on relations between terms created in the database and hearing aid type, and the preceding outcomes by audiologists [6].

These approaches will form the foundation for an audiology Decision Support System, where invisible records for patients would be given to the system, and the relative probability which the patient ought to be fitted with an ITE hearing aid rather than BTE hearing aid (or vice versa) would be returned.

REFERENCES

[1] Dr. Reza Sanati-Mehrizy, Utah Valley University, Jeffrey H Wright, Dr. Afsaneh Minaie, Utah Valley University, Dr. Ali Sanati-Mehrizy, Paymon Sanati-Mehrizy, Icahn School of Medicine at Mount Sinai, A Study of Application of Data Mining Algorithms in Healthcare Industry.

[2] R. Tamilselvi1, S. Kalaiselvi2, An Overview of Data Mining Techniques and Applications, International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064.

[3] Thair Nu Phyu, Survey of Classification Techniques in Data Mining, Proceedings of the International Multi Conference of Engineers and Computer Scientists 2009 Vol I, IMECS 2009, March 18 - 20, 2009, Hong Kong.

[4] J. Majumder, L. K. Sharma, Application of Data Mining Techniques to Audiometric Data among Professionals in India, Journal of Scientific Research & Reports, 3(23): 2960-2971, 2014; Article no. JSRR.2014.23.004, ISSN: 2320–0227.

[5] Boris Milovic*, Milan Milovic**, *Faculty of Economics Subotica, Novi Sad University, Serbia, **General Hospital Vrbas, Serbia, Prediction and Decision Making in Health Care using Data Mining, International Journal of Public Health Science (IJPHS), Vol. 1, No. 2, December 2012, pp. 69~78, ISSN: 2252-8806.

[6] Muhammad N Anwar, Michael P Oakes, Data mining of audiology patient records: factors influencing the choice of hearing aid type, ACM Fifth International Workshop on Data and Text Mining in Biomedical Informatics DTMBio 2011, Glasgow, UK. 24 October 2011.

[7] Michael Oakes, Shaun Cox, Stefan Wermter, School of Computing and Technology, David Goldman, Informatics Centre, Sir Tom Cowie, Campus at St. Peter's, University of Sunderland, SR6 0DD, England, Michael.Oakes@sunderland.ac.uk, Shaun.Cox@sunderland.ac.uk, Stefan.Wermter@sunderland.ac.uk, http://www.his.sunderland.ac.uk, Data Mining Audiology Records with the Chi-squared Test and Self-Organizing Maps.

[8] Muhammad Naveed Anwar, Knowledge Media Institute. The Open University. Milton Keynes, UK., Michael Philip Oakes, Faculty of Applied Sciences, University of Sunderland, Sunderland, UK, Decision Support System for the Selection of an ITE or a BTE Hearing Aid.

[9] Christo Panchev, Michael Oakes, Department of Computing, Engineering and Technology, University of Sunderland St. Peters Campus, Sunderland SR6 0RD, United Kingdom, christo.panchev@sunderland.ac.uk, michael.oakes@sunderland.ac.uk, Knowledge Media Institute, Muhammad Naveed Anwar, The Open University, Walton Hall, Milton Keynes MK7 6AA, United Kingdom, naveed.anwar@open.ac.uk, Hearing aid classification based on audiology data.

[10] John Keane, Principal Investigator, Michael Stone, Co-Investigator, University of Manchester, Manchester, United Kingdom, Lead Research Organization, Computer Science, Generating and using "big data" to identify hearing aid patterns of usage in order to optimise and personalise fitting.

[11] Kevin Munro, Ewing Professor of Audiology, kevin.j.munro@manchester.ac.uk, Michael Stone, Senior Research Fellow, michael.stone@manchester.ac.uk. University of Manchester, Manchester, Proposed audit of provision of hearing aids for mild/moderate hearing losses, M13 9PL, 25th November 2015.

[12] Clare Hiblin, National Direct Access Audiology PTL and WT Dataset Requirement for an Operational Information Standard, Information Standards Board Submission, V2.0 18 June 2007.

[13] J. Zhu, H. Zou, S. Rosset, T. Hastie, Multi-class AdaBoost, Statistics and Its Interface Volume 2 (2009) 349-360.

[14] Robert Burbidge, Bernard Buxton, Computer Science Dept., UCL, Gower Street, An Introduction to Support Vector Machines for Data Mining, WC1E 6BT, UK, 2001.

[15] Jason Brownlee, Logistic Regression Tutorial for Machine Learning, April 4, 2016 in Machine Learning Algorithms, https://machinelearningmastery.com/logistic-regression-tutorial-for-machine-learning/ .

[16] Janez Demsar, janez.demsar@fri.uni-lj.si, Orange: Data Mining Toolbox in Python, Journal of Machine Learning Research 14 (2013) 2349-2353.

[17] Bekkar, Dr.Hassiba Kheliouane Djemaa , Dr.Taklit Akrouf Alitouche, Evaluation Measures for Models Assessment over Imbalanced Data Sets, Journal of Information Engineering and Applications www.iiste.org ISSN 2224-5782.

[18] TJEN-SIEN LIM, A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms, Machine Learning, 40, Kluwer Academic Publishers, Boston. Manufactured in The Netherlands.