

Direct-to-Patient Survey for Diagnosis of Benign Paroxysmal Positional Vertigo

Heidi A. Richburg
Electrical and Computer Engineering
 Marquette University
 Milwaukee, USA
 heidi.hofschulz@marquette.edu

Richard J. Povinelli
Electrical and Computer Engineering
 Marquette University
 Milwaukee, USA
 richard.povinelli@marquette.edu

David R. Friedland
Otolaryngology and Communications
 Sciences
 Medical College of Wisconsin
 Milwaukee, USA
 dfriedland@mcw.edu

Abstract— Given the high incidence of dizziness and its frequent misdiagnosis, we aim to create a clinical support system to classify the presence or absence of benign paroxysmal positional vertigo with high accuracy and specificity. This paper describes a three-phase study currently underway for classification of benign paroxysmal positional vertigo, which includes diagnosis by a specialist in a clinical setting. Patient background information is collected by a survey on an Android tablet and machine learning techniques are applied for classification. Decision trees and wrappers are employed for their ability to provide information about the question set. One goal of the study is to attain an optimal question set. Each phase of the study presents a unique set and style of questions. Results achieved in the first two phases of the survey indicate that our approach using decision trees with filters or wrappers does a good job of identifying benign paroxysmal positional vertigo.

Keywords— *decision tree, classification, supervised learning, medical application*

I. INTRODUCTION

Dizziness is a common primary care complaint that remains evasive in specific diagnosis. Often a patient is misdiagnosed or referred to the incorrect specialist, prolonging the symptoms and associated dangers. Such misdiagnosis is costly, inefficient, and can cause patient harm. A known therapy for benign paroxysmal positional vertigo (BPPV), one of the most common causes of dizziness, can begin without delay if BPPV is diagnosed.

Hallmark symptoms of BPPV include vertigo spells that last seconds to one minute. These spells occur over weeks or months and may return after extended absence. The most common triggers include a change in the position of the head such as lying down in bed from a sitting or standing position, rolling over in bed, standing up, bending over, and tilting the head.

In an effort to identify BPPV among those who suffer from vestibular disorders, an electronic survey has been administered preceding a clinical visit with a vestibular therapist. The primary goals of the study are to apply machine learning methods to define the variables necessary for classification of BPPV and to create a clinical support system to classify the presence or absence of BPPV. Performance is measured with accuracy, sensitivity, and specificity. We strive to define a model with accuracy and specificity above 95%. Two iterations of the survey are completed with a third underway.

II. RELATED WORK

Several approaches have been taken to classify vestibular disorders, and BPPV specifically. A linear predictor was created by Friedland et al. [1] after identifying the most statistically relevant features ($p < 0.02$) from a 10-page patient intake questionnaire.

Multiple clinician-facing software systems have been developed to aid or train in diagnosis of vestibular disorders. “Vertigo” is a rules-engine expert system developed by Mira et al. [2] using patient background information and online learning to classify vertigo or dizziness from vestibular disorders. “Vertigo” asks questions until a diagnosis has been reached, and it terminates immediately if it determines the symptoms are not caused by a vestibular disorder. A second rules-engine expert system, “Carrusel”, has two stages [3]: the first stage uses patient history to narrow the diagnosis and recommend clinical tests for confirmation; the second stage incorporates clinical test results for a final diagnosis.

Two other studies incorporate clinical measures. “ONE” uses clinical data and an adaptive on-line format such that the software identifies and presents relevant questions as the survey progresses [4]. Kentala, Viikki, Juhola, and Pyykko used a large neurologic database to explore different modeling methods including decision trees, genetic algorithms, and neural networks [5-9]. We are focusing on decision trees for the readability they provide. With “EMBALANCE”, Exarchos et al. [10] examine all balance disorders and are working up to a dynamic, web-based early diagnostic system that would include clinical measures.

While “ONE” and “EMBALANCE” use clinical measures, our approach is based solely on survey questions. The patient enters their answers directly in our system, whereas all systems other than the linear predictor questionnaire involve a physician interfacing with software. Several studies have targeted BPPV specifically, including those by Chen et al. [11], Higashi-Shingai et al. [12], and Lapenna et al. [13]. In general, these studies tend to show that alternative processes such as patient questionnaires and predictive models can improve clinical diagnosis in a non-expert setting. However, they are limited by issues inherent to the use of static questionnaires.

Static questionnaires present a fixed set of questions to the patient. Limitations include missing data, question misinterpretation, and inability to confirm or better characterize specific variables [1],[13],[14]. For example, although a hallmark of BPPV is that spells of vertigo typically last less than a minute, asking a patient how long their dizziness lasts on a questionnaire is surprisingly inaccurate [13] [14]. Patients may have severe malaise for hours with an attack and fail to distinguish between the initial short spell of vertigo and the continuing hours of discomfort. Alternatively, individual episodes of BPPV can be recurrent over several days or weeks and then go away spontaneously, leading patients to interpret an episode as these longer time frames. While multiple questions could be asked regarding temporal patterns this leads to a lengthy questionnaire and increased likelihood of missing variables.

In an attempt to overcome limitations of static questionnaires we have initiated the use of an electronic format as a native Android application on an 8" Android tablet. This allows for branching toward pertinent questions and skipping others. It is critical to have accurate underlying predictive models for the questions needed to disambiguate, or classify, specific clinical diagnoses. This study represents the foundational steps in developing predictive models and in identifying decision trees for the diagnosis of BPPV.

III. EXPERIMENT SETUP

Approval for this study was obtained by the Institutional Review Board of the Medical College of Wisconsin, and all patients were consented prior to participation. Subjects were recruited from the vestibular therapy program within the Otolaryngology clinics at the Medical College of Wisconsin. The program is staffed by three certified vestibular therapists who see only vestibular patients. They work closely with neurotology service and have expertise in vestibular conditions including BPPV and confounding conditions.

Subject inclusion criteria included being a new adult patient being seen for initial vestibular therapy. Patients completed the survey at the time of their first session of vestibular therapy. A study coordinator instructed the patient on use of the tablet survey, which displayed one question at a time with the answer indicated by touching the appropriate choice. They could not advance to the next question until answering the current question. They could, however, go back and revise prior questions. At the end of each section a paragraph summarizing their answers was displayed and the patients had the option of repeating the section if incorrect.

At the time of survey administration, the study coordinator was blinded to the diagnosis and provided no guidance to the patients answering the questions. The study coordinator appended the patient's age to the nearest five years, gender, and diagnosis once available. Patient data were stored in SQLite databases as the survey was completed. Weka and MATLAB were used for data mining and analysis.

IV. SURVEY CREATION

A. Phase One Survey

The first phase of the survey began in May 2017. It was adapted from the 10 page, 162 variable paper-based survey used by Friedland et al. [1] and reduced to 84 questions. Seventy-two questions allowed 'Yes' or 'No' answers, and 12 of the questions were multiple choice. Minor branching was employed. For example, endorsing a single occurrence of dizziness eliminated specifying the frequency of spells. The questions were arranged in six sections with focus on: dizziness sensation; timing of attacks; triggers for attacks; migraine and headache history; otologic history; and medical history.

B. Phase Two Survey

The second phase of the survey consisted of a maximum of 68 questions and was administered beginning in November 2017. Of the 84 questions in phase one, 41 were removed since they were not identified as significant in the analysis. The order and grouping of questions also was adapted from the first phase. Some questions were added for clarification or based on variables collected by Siermala et al. [15].

To test whether the order in which temporal symptom questions are presented was significant, individual spell duration was shifted from before to after the overall symptom duration question in phase two. Several questions in phase one were identified as requiring clarification, which is accomplished with expansion and branching. For example, a phase one question:

- Do you have ringing or other noise in your ears (tinnitus)?

becomes two questions:

- Do you hear a ringing, buzzing, hissing, or other noise in your ears or head?
- Do you hear a ringing, buzzing, hissing, or other noise in one ear?

If 'Yes' is selected for either of these, two branch questions are asked:

- When did you first notice the ringing, buzzing, hissing, or other noise?
- How severe is the ringing, buzzing, hissing, or other noise?

These two branch questions were added to phase two of the survey to help disambiguate between BPPV and other diagnoses based on two variables used by Siermala et al. [15].

Phase two consisted of the same six focused sections as phase one. A supplemental section was added based on the two top performing decision trees from phase one to test how well these decision tree performed as standalone classifiers. Depending on responses, one to six questions were presented.

C. Phase Three Survey

The purpose of the phase three survey is two-fold. First, any unnecessary questions were eliminated in order to progress

toward the minimal question set. Second, an ensemble of weighted trees from the first two iterations was tested. Only the questions that appeared in a decision tree from phase one, phase two, the overlap of the two phases combined, or are critical from a clinician’s perspective were left in the question set. Branching was left intact such that only relevant questions were asked. The criteria for elimination was simple: if it did not appear on a tree or the clinician’s minimal list, then it was eliminated.

The questions were arranged in six sections as in the first two phases. Branching was employed as it was in phase two, though the supplemental section was excluded.

V. MACHINE LEARNING METHODS

This study included the identification of relevant questions for the patient-facing questionnaire, which meant a black box approach would not be feasible. Instead, it was desired to see the contents and structure of the model. All approaches chosen took this into account.

A. Decision Trees

Decision trees offer human readable results with the benefit of hierarchical information. The root of the tree contains the base question, and various child nodes of questions lead to the predicted result in leaves. This study used the J48 decision tree algorithm in Weka to classify each record as ‘Has BPPV’ or ‘No BPPV’.

B. Filters

In order to obtain a subset of the most relevant questions, attribute selection filters and wrappers were applied to the data. Filters rely on mathematical methods, and wrappers include the classification method. This study used the correlation attribute evaluator, which measures the Pearson’s correlation coefficient between each attribute and the class. The gain ratio attribute evaluator, which measures the gain ratio of an attribute with respect to the class, also was used. The wrapper subset evaluator was combined with the J48 decision tree and a greedy stepwise search for a third subset.

C. Direct Application of Two Decision Trees

A supplemental section was added to phase two of the survey to test the two top performing decision trees from phase one with a standalone dataset. The two trees had the same root and three common nodes. One tree had an additional question in one branch. This structure allowed for the trees to be nested. Questions were presented based on the resulting nested tree. The extra question was presented whenever that branch was followed, and then it was discarded in evaluating the smaller tree. If a leaf was reached, no more questions were asked. The questions were repeated from other sections of the questionnaire, but repeating them in the supplemental section allowed for the questioning to follow the decision tree structure for a direct test.

D. Linear Predictor Baseline

As a baseline, we compared our performance with that of the original paper-based questionnaire [1]. We also applied the linear predictor model directly based on the BPPV equation:

$$LP = -2.19 + 1.87x(\text{Lying down or rolling over}) + 0.92x(\text{Vertigo}) - 0.98x(\text{Symptoms for minutes to hours}) - 1.11x(\text{Symptoms for days}) - 1.84x(\text{Vertigo})x(\text{Symptoms for days to weeks}). \quad (1)$$

Direct application of (1) required a logical ‘OR’ operation on our data in some cases. For example, in phase two we split the lying down or rolling over in bed trigger question into two questions: one for a trigger of lying down in bed, and the other for a trigger of rolling over in bed. Similarly, the symptom duration questions had to be combined. The result of the linear predictor then was transformed into an estimated probability with equation:

$$Pr(BPPV) = \exp(LP)/[1 + \exp(LP)]. \quad (2)$$

E. Ensemble of Weighted Decision Trees

An ensemble learner often performs better than a single classifier. The third phase of the survey was designed to test an ensemble of the decision trees collected through phases one and two of the survey, including the combined dataset of overlapping questions. The trees can be tested individually, with majority vote ensemble, and as a weighted ensemble. Also, new trees can be discovered as new patient data are collected. Phase three of the survey will be administered starting in September 2018. Analysis will be flexible to allow integrating trees learned in phase three.

VI. RESULTS

Multiple methods were evaluated, and the methods with the highest consistent performance were presented and repeated for all phases. First, the J48 decision tree algorithm was applied to the unfiltered data set. Then, subsets were formed with the correlation attribute evaluator and the gain ratio attribute evaluator with thresholds of 0.25 and 0.1, respectively. Additionally, the wrapper subset was obtained. Finally the J48 decision tree algorithm was applied to the resulting subsets. The results are summarized in Table 1.

A. Phase One

The first phase of the survey was evaluated once data from the first 30 patients was available. It was repeated with 43, 58, 71, and 74 records. The tree changed at each analysis point, and then it remained the same from 71 to 74 records. We decided to proceed with phase two of the survey at this point.

The trees in Fig. 1 and Fig. 2 each have a root, three nodes, and five leaves, though the content differs between the two. It was no surprise that the trigger of lying down or rolling in bed appeared at the root of all three trees. However, none of the trees included a node relating to the length of symptoms, which is a hallmark of BPPV. Instead, the questions in the nodes served to disambiguate between BPPV and confounding conditions such as vestibular migraine. An accuracy of 0.92, sensitivity of 0.88, and specificity of 0.95 were achieved when the J48 decision tree was applied to the correlation attribute filter subset with data from 74 patients. The tree that was learned from this data is shown in Fig. 2.

TABLE I. SURVEY PERFORMANCE AND SIZE SUMMARY

Survey	Performance Per Survey			
	Accuracy	Sensitivity	Specificity	# of Questions
Paper Based	0.72	0.79	0.65	164
Phase 1	0.92	0.88	0.95	84
Phase 2	0.88	0.78	0.94	68
Supplement Tree 1	0.62	0.33	0.78	4
Supplement Tree 2	0.62	0.50	0.69	5
Combined Overlap	0.79	0.76	0.81	46

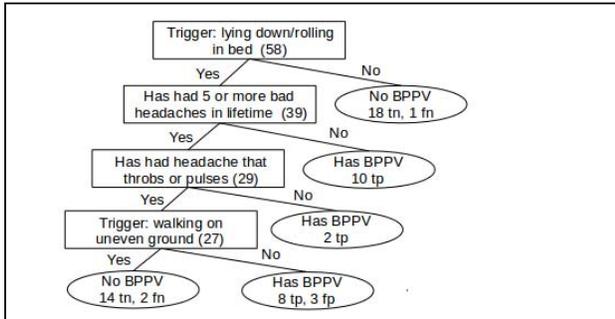


Fig. 1. Decision Tree with 58 Records in Phase One

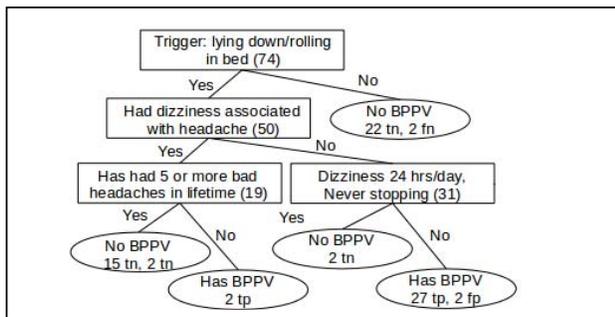


Fig. 2. Decision Tree with 74 Records in Phase One

B. Phase Two

Analysis was performed beginning with data from 30 patients and repeated per 10 additional records. The trees for the first few analysis points are shown in Fig. 3, Fig. 4, and Fig. 5.

The decision trees differed greatly from those in phase one. Many of the new questions appeared in the tree. Also, the length of symptoms, age, and gender appeared for the first time. Length of symptoms was for disambiguation rather than for identifying BPPV as the duration was hours but less than 12 hours. The disambiguation between BPPV and vestibular migraine was consistent with phase one only for the tree from the 30 patient records. Ear infections and ear pain were the other disambiguation points.

Performance for the tree in Fig. 5 was nearly as good as the trees in Fig. 1 or Fig. 2. It achieved an accuracy of 0.88, sensitivity of 0.78, and specificity of 0.94.

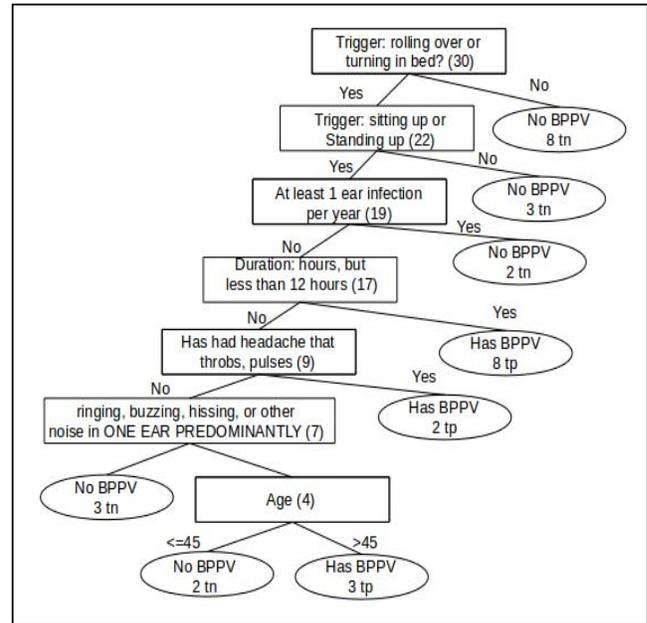


Fig. 3. Decision Tree with 30 Records in Phase Two

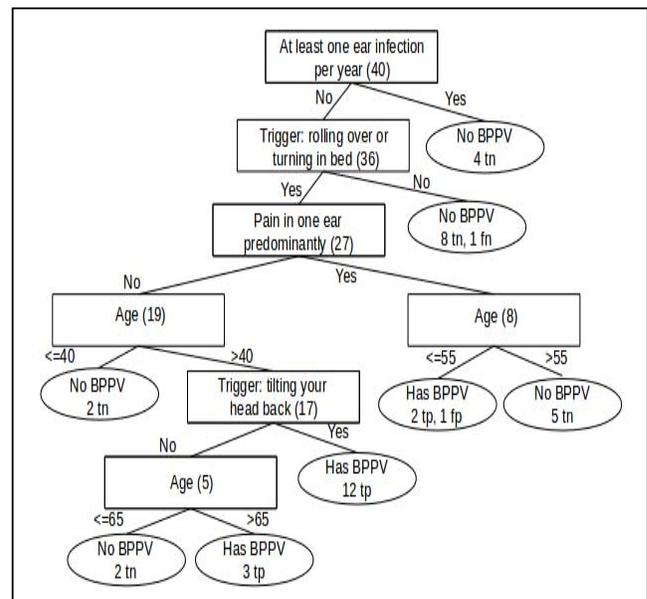


Fig. 4. Decision Tree with 40 Records in Phase Two

C. Combined Phases One and Two Overlap

At each analysis point for phase two, the analysis is repeated for the combined overlapping dataset of phases one and two. This provides a total of over 100 patient records beginning with the first analysis point. Phases one and two have a different number of questions and some question format differences. However with the inclusion of some logical 'OR' operations on a small set of phase two questions, 46 questions overlap.

The combined dataset tree structure remains stable when 30, 40, and 50 phase two patient records are used. It is interesting

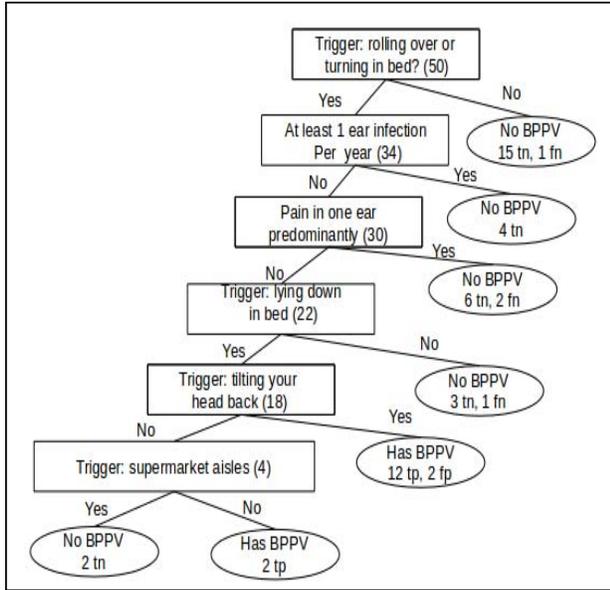


Fig. 5. Decision Tree with 50 Records in Phase Two

that it consists of only one root, one node, and three leaves. Again, the hallmark trigger appears at the root and a disambiguation question regarding migraines appears as its single child node. The subset produced by the correlation attribute filter with 30 and 40 phase two records did include the hallmark question of symptoms lasting seconds to one minute, however it did not become a node of the tree at any point. As simple as the tree in Fig. 6 may appear, it achieves 0.79 accuracy, 0.76 sensitivity, and 0.81 specificity.

D. Direct Application of Two Decision Trees

The performance of the nested decision trees was not as good as in phase one. The tree in Fig. 1, derived from results of 58 patients, fell from 0.90 to 0.62, from 0.87 to 0.33, and from 0.91 to 0.78 when applied to the phase two dataset. The tree in Fig. 2, derived from results of 74 patients, fell from 0.92 to 0.62, from 0.88 to 0.50, and from 0.95 to 0.69 when applied to phase two data. This difference in performance indicated that the direct application of two decision trees was not a sufficient model for consideration. Moving forward, the survey structure will continue to allow collection of sufficient data for a new tree to be learned.

It is unknown why the difference in performance for the same tree is so great on the phase two dataset. However, in comparing the responses in the supplemental section and the same questions in the main six survey sections, over half of the patients answered at least one question differently in the two places. Analysis was repeated using the responses from the main survey, and the results did not improve. It is impossible to determine the true answer to the question from the information we have. It also is impossible to gauge the patient's level of understanding of the question. Even though the survey is shorter than in phase one, it is possible patient engagement levels dropped before reaching the supplemental section.

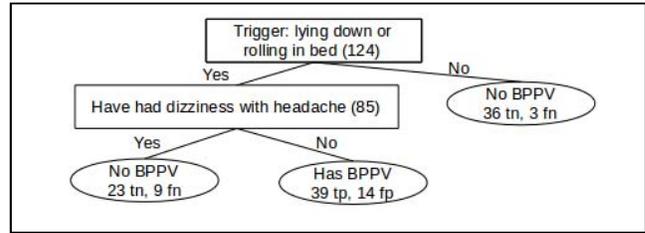


Fig. 6. Decision Tree with 124 Records in Overlapping Combined Phases

E. Linear Predictor Baseline

The linear predictor from the Medical College of Wisconsin paper-based survey achieved an accuracy, sensitivity, and specificity of 0.72, 0.79, and 0.65, respectively, with data from 381 patients. Phase one achieved 0.78, 0.55, 0.95 with 74 patients at its conclusion. Phase two achieved 0.62, 0.56, and 0.66 with 50 patient records. The combined overlap achieved 0.69, 0.71, and 0.67 with 124 records.

With one exception being the relatively high specificity attained in phase one, the results vary reasonably among the analysis points.

VII. CONCLUSION AND FUTURE WORK

Our method of using a decision tree to identify the presence or absence of BPPV performs reasonably well through phase two. The results offer an improvement over the paper-based questionnaire from [1]. Additionally, the fact that not just a single tree was learned from the data confirms that BPPV is a confounding diagnosis. The absence of the hallmark of duration in most trees indicates that the variable may not be extracted consistently from patients, as was found in [1][13][14].

A separate study will determine a standalone decision tree that can be used to most reliably extract the duration of each dizziness spell from the patient. Similarly, some education throughout the survey will help ensure that the patient understands and can differentiate among the survey definitions of the main symptoms: vertigo; imbalance; wooziness; swimming sensation; and faintness.

This BPPV study is a binary classification. Future studies can be multi-class, including other vestibular disorders or confounding conditions such as headaches. Since this study was biased by patients already identified for vestibular therapy, future work could include a wider patient population. The findings of this survey could be applied at a frontline provider, and the same survey would perhaps identify a different subset of relevant questions and decision trees.

ACKNOWLEDGMENT

We thank Sarah Mlezevia, Christy Erbe, and Rebecca Eernisse for their work coordinating with the clinic and patients.

REFERENCES

- [1] D. Friedland, S. Tarima, C. Erbe, and A. Miles, "Development of a statistical model for the prediction of common vestibular diagnoses". *JAMA Otolaryngol Head Neck Surg*, vol. 142, no. 4, pp. 351-6, 2016.

- [2] E. Mira, A. Buizza, G. Magenes, M. Manfrin, and R. Schmid, "Expert systems as a diagnostic aid in otoneurology", *ORL J Otorhinolaryngol Relat Spec.*, vol. 52, no. 2, pp. 96-103, 1990.
- [3] C. Gavilan, J. Gallego, and J. Gavilan, "'Carrusel': an expert system for vestibular diagnosis", *Acta Otolaryngol*, vol. 110, pp. 161-167, 1990.
- [4] E. Kentala, Y. Auramo, M. Juhola, I. and Pyykko, "Comparison between diagnoses of human experts and a neurotologic expert system", *Annals of Otology, Rhinology and Laryngology*, vol. 107, no. 2, pp. 135-140, 1998.
- [5] M. Juhola, J. Laurikkala, K. Viikki, E. Kentala, and I. Pyykko, "Classification of patients on the basis of otoneurological data by using Kohonen networks", *Acta Otolaryngol Suppl*, vol. 545, pp. 50-52, 2001.
- [6] E. Kentala, J. Laurikkala, I. Pyykko, and M. Juhola, "Discovering diagnostic rules from a neurotologic database with genetic algorithms", *Ann Otol Rhinol Laryngol*, vol. 108, no. 10, pp. 948-954, 1999.
- [7] E. Kentala, I. Pyykko, Y. Auramo, and M. Juhola, "Neural networks in neurotologic expert systems", *Acta Otolaryngol Suppl*, vol. 529, pp. 127-129, 1997.
- [8] K. Viikki, E. Kentala, M. Juhola, and I. Pyykko, "Decision tree induction in the diagnosis of otoneurological diseases", *Med Inform Internet Med*, vol. 24, no. 4, pp.277-289, 1999.
- [9] K. Viikki, E. Kentala, M. Juhola, I. Pyykko, and P. Honkavaara, "Generating decision trees from otoneurological data with a variable grouping method", *J Med Syst*, vol. 26, no. 5, pp.415-425, 2002.
- [10] T. Exarchos, C. Bellos, I. Bakola, D. Kikidis, A. Bibas, D. Koutsouris, and D. Fotiadis, "Management and modeling of balance disorders using decision support systems: the EMBALANCE project", *Advances in Experimental Medicine and Biology*, vol. 820, pp. 61-67, 2014.
- [11] W. Chen, et al., "Validation of 5-item and 2-item questionnaires in Chinese version of dizziness handicap inventory for screening objective benign paroxysmal positional vertigo", *Neurol Sci*, vol. 37, no. 8, pp. 1241-6, 2016.
- [12] K. Higashi-Shingai et al., "Diagnosis of the subtype and affected ear of benign paroxysmal positional vertigo using a questionnaire", *Acta Otolaryngol*, vol. 131, no. 12, pp. 1264-9, 2011.
- [13] R. Lapenna, et al., "Reliability of an anamnestic questionnaire for the diagnosis of benign paroxysmal positional vertigo in the elderly", *Aging Clin Exp Res*, vol. 28, no. 5, pp. 881-8, 2016.
- [14] D. Newman-Toker, "Symptoms and signs of neuro-otologic disorders", *Continuum, Minneap Minn*, vol. 18, no. 5 Neuro-otology, pp. 1016-1040, 2012.
- [15] M. Siermala, M. Juhola, J. Laurikkala, K. Iltanen, E. Kentala, I. Pyykko, "Evaluation and classification of otoneurological data with new data analysis methods based on machine learning", *Information Sciences*, vol. 177, pp. 1963-1976, 2007.