

BENIGN PAROXYSMAL POSITIONAL VERTIGO
PREDICTIVE DIAGNOSIS FROM
PATIENT-FACING SURVEY

by

Taiwo D. Fasae, B.S.

A Thesis submitted to the Faculty of the
Graduate School, Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Master of Science

Milwaukee, Wisconsin

December 2020

ABSTRACT

BENIGN PAROXYSMAL POSITIONAL VERTIGO PREDICTIVE DIAGNOSIS FROM PATIENT-FACING SURVEY

Taiwo D. Fasae, B. S.

Marquette University, 2020

Benign Paroxysmal Positional Vertigo (BPPV) is a leading cause of dizziness and imbalance that is responsible for one-third of fall incidents. Diagnosis, however, is ridden with uncertainties and errors. This thesis explores various techniques for BPPV predictive diagnosis from a survey study and proposes measures for predictive performance improvement. Patient-facing surveys are established ways of acquiring medical history in clinical settings and, as this thesis demonstrates, are capable of conveying patterns distinguishable for accurate diagnosis.

This work begins by discussing BPPV and vestibular disorders in general, and the risks associated with misdiagnosis or elusive diagnosis. Innovative efforts by medical professionals in vestibular therapy for handling the intricacies of diagnosis and clinical protocols are also explained. To predict BPPV successfully, there are distinguishing marks present in a patient's dizziness episodic history including the frequency and duration of episodes, the specific nature of the dizziness, and the positional trigger. Given these indicators for predicting BPPV, we develop a number of statistical models on a dataset of survey responses acquired from a clinical cohort study.

Next, the thesis establishes a connection between the performance limits of the machine learning methods, and the existence of incorrect answers to the survey prompts. By demonstrating that question misinterpretation and ambiguities exist in the cohort study, we show that certain data quality improvement measures have significant influence on classification performance.

ACKNOWLEDGEMENTS

Taiwo D. Fasae, B.S.

I would like to thank the KID Lab of the Electrical and Computer Engineering department at Marquette University for providing me the platform and support which enabled me to undertake this research work.

I am especially grateful to my advisor, Dr. Richard Povinelli for providing me with this unique opportunity, providing insights and direction, and for believing in my ability. Special thanks to Dr. George Corliss for his guidance, input, and mentorship throughout this work, and to Dr. Ronald Brown for his feedback and support all through. Many thanks to Dr. David Friedland of the Medical College of Wisconsin for spearheading the project from which this work came to life. To past students, Heidi Richburg and Hong Ding, who have laid the groundwork that made this thesis possible, I owe my gratitude.

Special thanks to the members of the Knowledge and Information Discovery Lab. Notably, Colin Quin for his willingness to read and offer suggestions to improve this publication. To Masabho Milali for providing encouragement and support throughout this work. Other members of the KID lab and Marquette Energy Analytics, including but not limited to David Kaftan and Zachary Nordgren, for their support all the way.

Many thanks to my thesis committee in the persons of Dr. Richard Povinelli, Dr. George Corliss, Dr. Ronald Brown, and Dr. Dong Hye Ye for their valuable advice, comments, and scrutiny.

This work is dedicated to my parents, Drs. Peter and Kehinde Fasae, and siblings Kenny, Kemi and Ayo.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i
TABLE OF CONTENTS.....	ii
LIST OF TABLES	iv
LIST OF FIGURES	vi
CHAPTER 1 BENIGN PAROXYSMAL POSITIONAL VERTIGO.....	1
1.1 Contribution.....	1
1.2 Motivation	2
1.3 Vestibular Disorders and BPPV	3
1.4 BPPV Diagnosis and Therapy	3
1.5 BPPV Indicators and Symptoms	4
1.6 Importance of Accurate BPPV Diagnosis	4
1.7 Why BPPV Predictive Diagnosis?	5
1.8 Organization of Thesis.....	6
CHAPTER 2 PREDICTIVE DIAGNOSIS OF BENIGN PAROXYSMAL POSITIONAL VERTIGO	7
2.1 Benign Paroxysmal Positional Vertigo.....	7
2.2 Prior Work	9
2.3 Machine Learning Models.....	11
2.3.1 Logistic Regression Classifier	12
2.3.2 Naïve Bayes Classifier.....	13

2.3.3	Artificial Neural Network	14
2.3.4	Decision Tree Classifier.....	15
2.4	Summary.....	17
CHAPTER 3 PATIENT-FACING SURVEY PROCESS		18
3.1	The Survey Process	18
3.2	Survey Section Areas.....	19
3.3	Medical Data and Health Information Privacy	19
3.4	Survey Overview	20
3.5	Survey Statistics	21
3.6	Diagnosis Statistics.....	23
3.7	Summary.....	24
CHAPTER 4 BPPV PREDICTION RESULTS		25
4.1	Performance Metrics.....	25
4.2	Logistic Regression	25
4.3	Naïve Bayes algorithm	28
4.4	Artificial Neural Network.....	30
4.4.1	Autoencoder for Dimensionality Reduction	30
4.4.2	ANN Classifier	32
4.5	Decision Trees	34
4.5.1	Correlation Attribute Evaluation.....	34
4.5.2	Model Performance on Single Questions.....	36
4.6	Opportunities for Improvement	37

4.7	Summary.....	38
CHAPTER 5 IMPROVING THE USER SURVEY EXPERIENCE		39
5.1	Motivation	39
5.2	Importance of Accurate Survey Responses	40
5.3	Average Response Time Vs. Misclassifications.....	40
5.4	Inaccuracies in Reporting of Dizziness Episodes Duration.....	42
5.5	Symptoms Misinterpretation	43
5.6	Performance Improvement Measures and Simulation.....	44
5.6.1	The “Vertigo” Symptom Question.....	44
5.6.2	The “Lying Down/Rolling Over” Trigger prompt.....	45
5.6.3	The Dizziness Spell Timing Prompt.....	46
5.6.4	All Three Assumptions	47
5.7	Survey User Education	49
5.7.1	In-Survey Tutorial	50
5.7.2	Interface Redesign	50
5.7.3	Language Simplification.....	50
5.7.4	Response Consistency Check	51
5.8	Summary.....	51
CHAPTER 6 CONCLUSIONS AND RECOMMENDATIONS		52
6.1	Research Contributions.....	52
6.2	Recommendations	52
Bibliography		54
APPENDIX I Selected Screenshots from the BPPV Survey.....		59

LIST OF TABLES

TABLE 3-1: SURVEY QUESTION DETAILS	20
TABLE 3-2: SURVEY QUESTION DETAILS (CONTD.)	21
TABLE 3-3: ATTRIBUTES CORRELATION COEFFICIENT AGAINST BPPV DIAGNOSIS. SIGNIFICANT CORRELATION VALUES ARE IN BOLD LETTERS	22
TABLE 3-4: (CONTD.) ATTRIBUTES CORRELATION COEFFICIENT AGAINST BPPV DIAGNOSIS. SIGNIFICANT CORRELATION VALUES ARE IN BOLD LETTERS	23
TABLE 4-1: NAÏVE BAYES CROSS-VALIDATED MODEL PERFORMANCE WITH/WITHOUT FEATURE REDUCTION	28
TABLE 4-2: DETAILS OF THE AUTOENCODER CONFIGURATIONS WITH THE LEAST RECONSTRUCTION ERRORS	32
TABLE 4-3: 10-FOLD CROSS VALIDATION RESULTS OF VARIOUS FEED-FORWARD ANN CONFIGURATIONS ON THE DIFFERENT AUTOENCODER OUTPUTS IN THE PREVIOUS TABLE (HIGHEST VALUES FOR AUC AND ACCURACY ARE HIGHLIGHTED IN RED)	33
TABLE 5-1: TREE MODEL PERFORMANCE ON THE ADJUSTED DATASET	48

LIST OF FIGURES

FIGURE 2-1: BENIGN PAROXYSMAL POSITIONAL VERTIGO AND THE CANALITH REPOSITIONING PROCEDURE ©MAYO [18] WITH PERMISSION.....	8
FIGURE 2-2: ARTIFICIAL NEURAL NETWORK ARCHITECTURE	14
FIGURE 2-3: A SAMPLE DECISION TREE FOR BPPV DIAGNOSIS	16
FIGURE 4-1: RECEIVER OPERATING CHARACTERISTICS OF THE LOGISTIC REGRESSION MODEL, AUC=0.78.....	27
FIGURE 4-2: A 13-NODE HIDDEN-LAYERED AUTOENCODER PICTORIAL REPRESENTATION	31
FIGURE 4-3: A THREE-LAYERED FEED-FORWARD ANN CLASSIFIER.....	33
FIGURE 4-4: DECISION TREE ON OVERALL DATASET	35
FIGURE 4-5: DECISION TREE ON "COLLAPSED SENSATION" QUESTION	36
FIGURE 4-6: DECISION TREE ON "LYING DOWN/ROLLING OVER"	37
FIGURE 5-1: HISTOGRAM OF AVERAGE TIME SPENT PER QUESTION.....	41
FIGURE 5-2: PREDICTION ACCURACY BY AVERAGE TIME SPENT PER QUESTION.....	42
FIGURE 5-3: PERCENTAGE OF "YES" RESPONSES TO THE DURATION QUESTIONS.....	43
FIGURE 5-4: TREE MODEL OBTAINED FROM "VERTIGO" ADJUSTED DATA	45
FIGURE 5-5: TREE MODEL RESULT FROM ADJUSTED "TRIGGER: LYING DOWN" PROMPT.....	46
FIGURE 5-6: TREE MODEL RESULT FROM "DURATION: BRIEF SPELL" PROMPT	47
FIGURE 5-7: TREE MODEL RESULT OF "ALL 3 ATTRIBUTES" ADJUSTMENT	48

CHAPTER 1 BENIGN PAROXYSMAL POSITIONAL VERTIGO

Dizziness and vestibular problems are common complaints for which patients seek medical assistance; about 40% of the United States population will experience some form of vestibular distress in their lifetime [1]. Individuals with vestibular disfunctions experience interruption in their daily activities and in their social and work life. A dizziness study reports 27% of subjects with dizziness changed jobs, 50% had reduced efficiency at work, and 57% had a disruption in their social life [2]. Another study found a 12-fold increase in the odds of falling – a higher risk among older individuals [3] – and 10% of falls result in major and severe injuries including death [4]. Unfortunately, misdiagnosis of vestibular disorders is common, inefficient, and ridden with patient frustration and dissatisfaction [5]. Moreover, in an era where in-person clinical appointments are less appealing, efficient and accurate diagnostic procedures are necessary. In this thesis, we turn to machine learning for predictive diagnosis of Benign Paroxysmal Positional Vertigo (BPPV) – a leading cause of dizziness.

1.1 Contribution

We present the results of applying four unique machine learning algorithms to questionnaire data acquired from a clinical cohort study for BPPV. Logistic regression, Naïve Bayes, Artificial Neural Network, and Decision Trees algorithms are assessed on the data that captures the nature, timing, and triggers of dizziness of 397 patients. We show that the prediction performance of the models is limited by inaccuracies in the

responses to certain crucial questions, and we demonstrate the existence of these inaccuracies. For example, BPPV is triggered by head position changes and lasts for 30 to 60 seconds, as such, the data is expected to reflect this. Lastly, we propose certain measures to the patient-facing survey that promise better data quality and consequently, better predictive performance. Measures such as the introduction of in-survey symptoms education will improve the patient's understanding of the prompts and consequently, more accurate responses.

1.2 Motivation

BPPV, like many causes of dizziness and imbalance, can lead to stumbling, motion sickness or inability to do simple balance activities. Imbalance can devolve into falling causing severe injuries and fatalities, especially in geriatric patients: the death rate from unintentional falls among aged 65 and older persons in the US have increased since 1993, Center for Disease Control [6]. Misdiagnosis, unfortunately, is common [7], inefficient and high-risk. For instance, BPPV misdiagnosis rates in emergency departments and primary care offices is about 80% [5]. Furthermore, the clinical procedures to differentiate BPPV from other origins of dizziness, e.g., stroke, are lengthy and dissatisfying [5], and place the patient at great risk of denied emergent care. For these reasons, computation techniques have been advocated to improve the diagnosis of vestibular disorders [8]–[10]. Pattern recognition has shown potential in discovering patterns in medical history and symptoms [8], [10]–[12], and we assess four machine learning algorithms on the data in this thesis.

1.3 Vestibular Disorders and BPPV

A vestibular disorder occurs when there is a malfunctioning of the human vestibular system [13]. This is common when certain structures of the human ear that are responsible for balance and eye movements are defective or damaged. Damage caused by diseases, ear infections, side effects of medication, aging, or injury to any of these delicate structures often result in abnormalities in balance, and present symptoms such as vertigo, light headedness, wooziness, or a swimming sensation or a combination of them [1]. Even mild symptoms such as erratic sensitivity to light can be unsettling and pose a serious danger to the patient. Therefore, vestibular disorders can significantly upset a person's daily routine and cause psychological and emotional hardship.

BPPV is a brief, intense episode of vertigo that occurs when loose otoconia (bio-particles in the inner ear) emerge in one of the semicircular canals of the inner ear important for vestibular function [1]. This spinning sensation is mostly triggered by sudden changes in head position and typically lasts between seconds and a minute.

1.4 BPPV Diagnosis and Therapy

BPPV diagnosis can be complex, lengthy and a resource-intensive procedure [5]. When BPPV emerges as a potential prognosis, a patient is examined to identify signs and symptoms necessary to establish the need for further procedures. However, there are situations where these measures are inconvenient or impossible to carry out, e.g., a patient with a neck pathology. In other cases, neurological tests are carried out before further examination or therapy. These exert a burden on clinical resources and the clinical

support system. These limitations have inspired care providers to find alternative ways to diagnosis. Machine learning is an established approach for use in the diagnosis of vestibular disorders [8], [10], [11].

1.5 BPPV Indicators and Symptoms

Symptoms of vestibular disorders may be mild or severe, may last only seconds or minutes, or may occur persistently. Ear infections, head trauma, exposure to pressure changes, long-term medications, migraine, stroke are known causes of vestibular malfunctions [13]. For BPPV, there are more specific hallmark symptoms. BPPV patients experience vertigo in brief episodes lasting less than one minute; usually triggered by changes in head position; and return to normal between episodes [14]. These indicators are critical to the success of the predictive models under consideration in this study.

1.6 Importance of Accurate BPPV Diagnosis

Vestibular disorders manifest as unsteadiness and dizziness. Thus, patients with vestibular difficulties are at a greater risk of falling. They may stagger when they try to walk, stand, or perform normal day-to-day activities. In worse cases, some patients experience blurred vision, nausea, or appear confused or disoriented. A misdiagnosis, which is unfortunately common, may mean that a patient's health continues to deteriorate, or an otherwise healthy patient is made to go through unnecessary and difficult procedures.

Dizziness is a common complaint in emergency rooms and a symptom of a host of diseases some of which are life-threatening and emergent, e.g., stroke. As the safest course of action, expensive tests and procedures such as CT scans are usually conducted to eliminate the chances of these emergent conditions [15]. Many such cases get misdiagnosed multiple times, even when eventually diagnosed correctly.

Moreover, a vertigo incidence resulting from BPPV has a 50% recurrence rate [16], which means that patients could grapple with a poor quality of life [17]. Thus, the cost of misdiagnosis of BPPV is large both in the quality of life of affected patients and in clinical resource allocation.

1.7 Why BPPV Predictive Diagnosis?

As a result of the undesirable outcomes and frustrations in the diagnosis of vestibular disorders, machine learning can prove beneficial in discovering underlying patterns in patient medical data. For an accurate diagnosis, the right patterns and indicators can be identified and learned for accurate prediction. Machine learning techniques such as Artificial Neural Network, Logistic Regression and Support Vector Machines are algorithms with a proven track record in pattern recognition and classification. The Decision Tree algorithm also is explored in this work because it is suited for modelling medical decisions and human interpretability. Using machine learning techniques, a pre-encounter vestibular assessment survey may be sufficient for a successful BPPV diagnosis, and potentially improve the clinical diagnostic process. In an era of patient-centered care and remote monitoring, innovative ways of providing better and smarter

clinical support systems is gaining momentum [8], [10], [11], and artificial intelligence is proving more valuable than ever in the clinical diagnosis process.

1.8 Organization of Thesis

This work focuses on improving the accuracy of prediction models for BPPV. Chapter 2 discusses prior work in BPPV predictive diagnosis. The methods used in this work are discussed and explained.

In Chapter 3, we discuss the survey data-gathering process beginning from when a patient visits the clinic to the follow-on session with a specialist. The survey sections and question formats are described, and the resulting dataset attributes are summarized.

Chapter 4 discusses the prediction models explored on the survey data, with extra attention on decision tree models. We compare performances across the four models and explain the reasons for the limits in prediction performance.

Chapter 5 discusses the realities and expectations of the survey outcome, and we show that prediction performance limits are influenced by the correctness of the survey responses. We also show why certain data quality improvement measures promise a prediction model with better performance than the current prediction models obtained in this work.

In the concluding chapter, we present an overview of the research contributions and propose future work and possible lines of research continuation in predictive BPPV diagnosis.

CHAPTER 2 PREDICTIVE DIAGNOSIS OF BENIGN PAROXYSMAL POSITIONAL VERTIGO

This chapter discusses BPPV along with the associated treatment and therapy procedures. The statistical models and prediction algorithms employed are presented. Popular algorithms, such as Logistic Regression, Bayesian Models, Artificial Neural Network, and Decision Trees, are explored.

2.1 Benign Paroxysmal Positional Vertigo

BPPV is a brief and intense episode of vertigo triggered by a change in the position of the head that occurs when loose crystals or particles tumble into one of the semicircular canals of the inner ear [1]. As Figure 2-1 illustrates, these semicircular canals contain fluid, hairlike sensors, and crystals (otoconia) that monitor and maintain the body's sensitivity to gravity. The unwanted presence of otoconia in the semicircular canal makes a person abnormally or overly sensitive to changes in head position. This spinning sensation triggered by sudden head position changes is named Benign Paroxysmal Positional Vertigo. The vertigo is benign because the sensation is rarely serious, paroxysmal because the sensation happens quickly, usually from 30 to 60 seconds, and positional because the sensation is triggered by changes in head position.

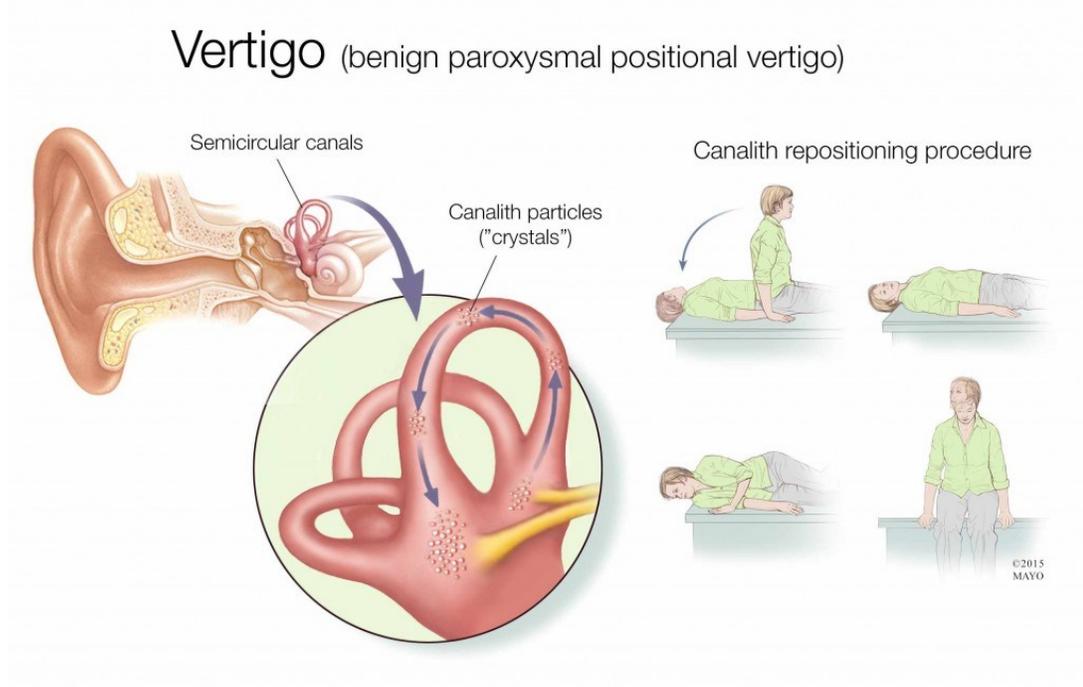


Figure 2-1: Benign Paroxysmal Positional Vertigo and the Canalith repositioning procedure ©MAYO [18] with permission

BPPV, although benign, is a leading cause of dizziness, hence puts the patient at risk of falling. About 35% of Americans above the age of 40 have experienced one form of vestibular dysfunction, and 80% of those above the age of 65 have experienced dizziness [3]. As such, the risk of having imbalance complications increases with age. For instance, dizziness is linked to fall incidents in the elderly [19].

Because BPPV treatment and therapy requires lengthy procedures and substantial clinical resources, suspected patients usually are either unable to get immediate care or find the clinical process tiresome and dissatisfying [5]. A significant cause of misdiagnosis is the awareness among care providers that dizziness is a symptom of a

multitude of illnesses. As such, there is a knowledge gap on what information is relevant and critical to differentiating BPPV [20][21].

BPPV in some cases goes away on its own, but to speed the recovery process, a well-organized vestibular therapy plan is carried out and taught to patients to perform at home when needed. This may include series of movements such as the Canalith repositioning method illustrated in Figure 2-1. The Canalith repositioning involves a simple reposition of the head to move the dislodged bio-crystals into the vestibule, the location where the particles get absorbed [22]. Sometimes, surgery is required. A bone plug is used to seal the deficient part of the ear to prevent the crystals from moving.

2.2 Prior Work

Vestibular diagnoses have symptoms that overlap between different disorders (each pathophysiology requires unique treatments) and thus, a confounding diagnosis problem. With the patient-intake questionnaire method, relevant patient history can be obtained in a structured format, and machine learning can be used to recognize patterns crucial to BPPV diagnosis.

Data analysis in medical diagnosis has seen great interest in vestibular disorders and BPPV in particular [10]–[12]. Despite the popularity of BPPV among dizziness complaints, cases in the emergency department (ED) have a poor diagnosis record [21]. Because dizziness is both a symptom of BPPV and some serious medical conditions, e.g., stroke, emergency room physicians usually do not have the time and training to diagnose all medical conditions, and therefore play it safe by performing expensive tests and scans to eliminate the serious threats [15]. Overreliance on diagnostic scans and patient's

description of symptoms is another known cause of BPPV misdiagnosis [15], [23]. To improve vestibular diagnostic measures, alternative and complementary processes such as algorithm diagnostics, intake questionnaires, and predictive models have been employed [24-30].

VERTIGO, a system developed by Mira et al. [24], uses patient background information and online learning to distinguish vertigo from vestibular disorders by asking questions until a diagnosis has been reached. A similar tool, CARRUSEL [25] narrows the diagnosis from patient history and recommends further clinical tests which it incorporates for a final diagnosis. These methods, however, require clinical tests and measures for effective output. In addition, structured patient information, which is a requisite for these algorithms, is obtained through very inefficient and rigorous procedures of processing triage notes and voice recordings. As an alternative approach, a direct-to-patient elicitation of information has been proposed for obtaining patient history data [26].

Several studies have targeted BPPV diagnosis with the use of survey questions. In a BPPV questionnaire of six questions [27], Kim H-J et al. achieved 80% precision in classifying BPPV and the specific subtype of the diagnosis. Similarly, Higashi-Shingai et al. [26] found a four-question survey accurate for 80% of the subjects in the study. The survey-questionnaire approach is very simple and easy to administer. But, in many of these settings, the questions are static and non-customized to every patient's peculiar condition. In some cases where the questions are presented in branching order or adaptive mode, some questions – potentially relevant to the diagnosis – are ignored. For instance, indicators such as the duration of a vertigo spell or the frequency of migraines are subtle

details that cannot be diagnosed efficiently by a fixed set of questions. Notwithstanding, the success of these methods demonstrates that prediction algorithms can find and recognize patterns in patient data and history relevant to BPPV.

Significant efforts in BPPV prediction from survey data have been relatively fruitful. Friedland et al. [28] showed that four predictors from a patient-intake pre-encounter questionnaire could predict BPPV using logistic regression with sensitivity and specificity of 79% and 65%, respectively. When the same questionnaire and statistical model was applied to 200 medical records [29], sensitivity and specificity of 76% and 100%, respectively, were obtained. Richburg et al. [30] presented a decision tree analysis on a survey data of 381 patients and obtained 72% accuracy. The authors deduce that BPPV is a confounding diagnosis, and the results explain why the misdiagnosis rates in practice are undesirable. There are, however, certain classic symptoms of BPPV that should be present in the model (by nature of the indicators of BPPV) but were absent. This demonstrates the existence of certain hurdles in patient self-reported symptoms, e.g., question misinterpretation in BPPV diagnosis.

2.3 Machine Learning Models

We applied four prediction techniques to our data to derive models capable of predicting BPPV diagnosis in a fully supervised classification fashion. This section provides the necessary background of logistic regression, Naïve Bayes classifier, artificial neural networks, and decision trees to the BPPV survey data.

2.3.1 Logistic Regression Classifier

Logistic regression is a statistical algorithm that models the relationship between a categorical dependent variable and a number of independent variables using regression analysis [31]. This is based on the assumption that there exists a linear combination of independent variables that is logistically related to a category output. Logistic regression is similar to linear regression in terms of deriving a linear combination of a set of predictors, except for having an extra layer of a sigmoid (or similar) function that maps the result to $[0,1]$. In prediction, the class with the highest value is designated as the predicted class.

In many medical diagnosis problems where there is the need to classify a patient into two categories or possible outcomes (sick or healthy, diagnosed or not, etc.), the probability of belonging to one outcome complements that of the other. This scenario is known as a binary logistic regression problem. In the training process, the cross-entropy and the gradient descent algorithm are generally used as the cost function and optimization method, respectively. To predict to which class a patient belongs, the class (diagnosis) with the higher probability is selected. In a binary classification case (as in this work), the 50% probability threshold is commonly adopted. For a probability output value above 50%, the record is assigned to one class, otherwise the record is assigned to the other class.

Logistic regression has been successful in predicting BPPV using four variables of patient intake questionnaires. Britt et al. [28] demonstrates an outcome of sensitivity and specificity of 0.79 and 0.65, respectively. But in this work, we explore the logistic

regression model for predicting BPPV and contrast the results with other models discussed and those from the literature. These are presented in Section 4.2.

2.3.2 Naïve Bayes Classifier

The Naïve Bayes algorithm is a probabilistic estimation method based on the Bayes theorem. The “naïve” assumption is that each predictor is independent of the others. This is known as the Naïve Bayes rule because of the seemingly naïve assumption of independence amongst predictors. In mathematical terms, if X and y represent the feature vector and class, respectively, according to Bayes’ theorem,

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}.$$

2-1

Therefore, by assuming the predictors are independent of each other, and with equal contribution to the class, $P(X|y) = P(x_1|y)P(x_2|y) \dots P(x_n|y)$ becomes

$$P(y|x_1, \dots, x_n) \propto P(y) \prod_{i=1}^n P(x_i|y).$$

2-2

The input probabilities are obtained either from prevalence given in literature or estimated from the dataset. For prediction, the class with the maximum likelihood is selected as the predicted class.

Naïve Bayes classifiers have been used in predictive medical diagnosis to some level of success. In [32], medical records are used to successfully construct an ontology

and train Naïve Bayes classifier to evaluate the probability of 31 different diseases. With these in mind, we explore the Naïve Bayes classifier on our survey data and present the results in Section 4.3.

2.3.3 Artificial Neural Network

Artificial Neural Networks (ANN) are a class of machine learning algorithms made up of an interconnection of nodes that mimic a network of neurons as in the human nervous system. The architecture comprises of two or more layers connected in a forward propagation paradigm. By applying certain activation functions to the hidden layers or output layer or both, very complex relationships and underlying patterns can be learned and recognized. Figure 2-2 shows a simple ANN architecture with three input nodes, one hidden layer of 4 nodes, and two output nodes.

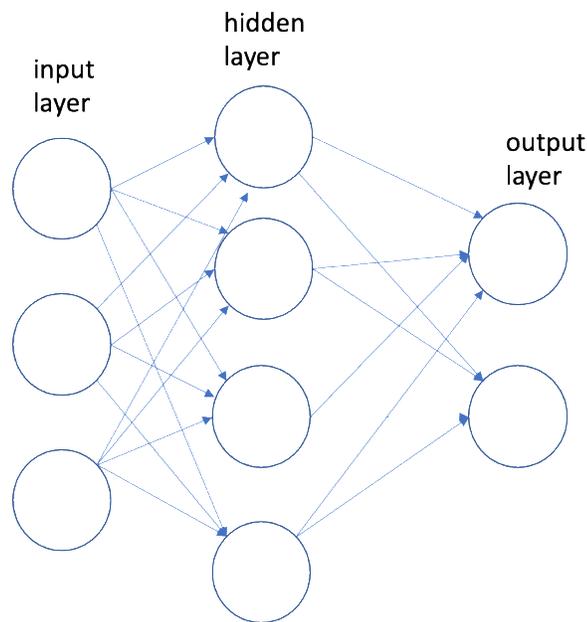


Figure 2-2: Artificial Neural Network architecture

ANNs are characterized by containing weights along the paths between neurons and tuned by a learning algorithm to improve the model. The weights are initialized prior to training. During training, data goes through each node in the input layer through activation functions towards those in output layer in a feed-forward fashion. The most commonly used activation function is the sigmoid function. An optimization technique is used for determining the best values for all of the tunable model parameters. The cost function is used to learn the optimal solution to the problem solved. The problem is solved when the output of the cost function is minimized or the target number of iterations is reached, or the optimization function is converged.

ANNs are well suited for classification tasks where the contents and structure of the model is relatively unimportant. Such black box models, while may be undesirable in medical diagnosis due to the lack of human interpretation of the model structure, can provide impressive prediction performance [11].

2.3.4 Decision Tree Classifier

A Decision Tree is a structure in which each node represents a decision, with each outcome (branch) leading to another decision until the bottom of the tree is reached. An example of a decision tree for a BPPV prediction is shown in Figure 2-3. For example, given a set of responses of a patient and the tree in Figure 2-3, we begin the classification process at the top node. If the patient's response to "Trigger: lying down?" is a No, then the classification output is the "No BPPV" diagnosis case. Otherwise, the output depends on the response to the "Vertigo?" question. If No, then the output is "No BPPV", otherwise, the decision depends on the response to the "Last 1 min?" question.

The decision tree classifier builds classification by decomposing the data set into smaller and smaller subsets while an associated decision tree is developed incrementally. The final result is a tree with decision nodes (node with branches) and leaf nodes (without branches).

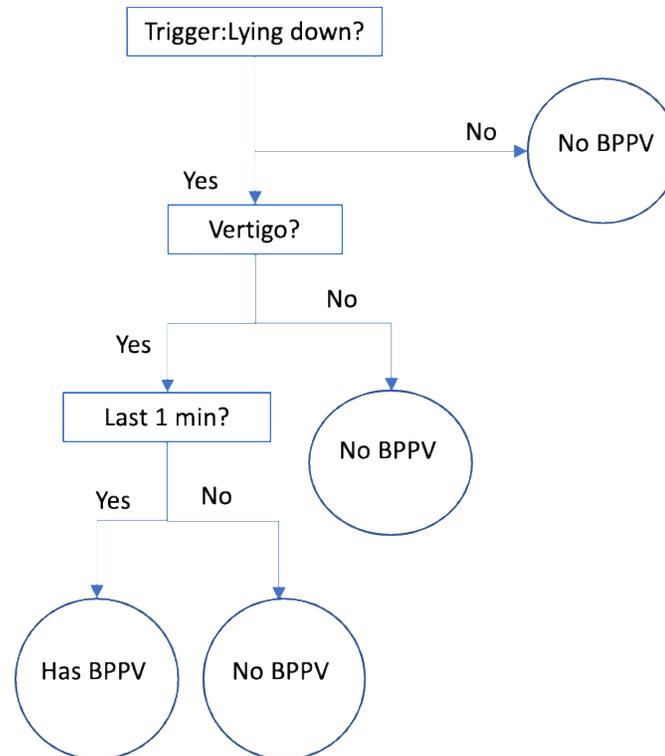


Figure 2-3: A sample Decision Tree for BPPV diagnosis

There are two major steps (splitting and pruning) involved in building a decision tree. Splitting is a recurrent process of partitioning a subset into smaller splits or subsets. The strategy of splitting is to partition the data along the best line of split, i.e., the attribute that best splits the data according to some metric such as information gain. Information gain is a measure of how much information is gained after a dataset is split on an attribute. The attribute with the highest information gain is established as the best

attribute with which to split the subset and is designated as the decision node at that point in the splitting process. Entropy, a calculation of the homogeneity of a sample, is used to measure the information content exhibited by an attribute. If a sample is completely homogenous, there is no information, and the entropy is zero. But, if the sample is equally divided, it has entropy of one. The pruning process involves reducing the size of a tree by cutting off subtrees that either diminish performance or do not improve it at all.

Because Decision Trees are made up of branching and logic decisions, they have the advantage of human interpretability. This unique feature of a decision tree model is highly attractive for its use in medical diagnosis and clinical decision support systems.

2.4 Summary

Machine learning has notable applications in vestibular assessments and medicine. The nature of vestibular complaints and the myriads of illnesses causing dizziness make diagnosis hard and frustrating. From the literature, we reviewed existing applications of the Logistic Regression, Naïve Bayes, Artificial Neural Networks and Decision Trees classifiers in vestibular disorders and BPPV diagnosis.

CHAPTER 3 PATIENT-FACING SURVEY PROCESS

The data used in this work was obtained from a series of phases of clinical surveys aimed at guiding patients to provide accurate description of symptoms and the nature of their dizziness [30]. This chapter presents a statistical overview of the data and describes the survey process.

3.1 The Survey Process

Subjects in this work are recruited from the vestibular therapy program within the Medical College of Wisconsin. The survey is taken by adult patients visiting first time for initial evaluation for vestibular therapy.

Prior to a patient's first encounter with a specialist, the survey is provided as an Android application on a 10-inch electronic tablet. The patient is expected to provide answers to the questions on the survey with minimal help of clinical support staff. The survey was designed such that only one question/prompt is given to the patient at a time. Each prompt requires a response, and the respondent cannot proceed without providing an answer, although an earlier response can be adjusted.

After taking the survey, the resulting data is accessed through the cloud for compilation and data analysis. Each survey record is labeled with the diagnosis obtained from a specialist with a knowledge of whether the patient has BPPV or not. The cohort study is an ongoing long-term research effort with preliminary reports in [30].

3.2 Survey Section Areas

Given that a patient's complete and accurate history is crucial to the success of BPPV diagnosis [20], the survey is designed to obtain such relevant information. Thus, an answer to every question on the survey is required, and there are no missing or incomplete records, although the accuracy of some answers may be suspect.

The survey is grouped into segments corresponding to a class of symptoms including:

1. The nature of the dizziness experienced; a series of Yes/No questions to describe the onset, duration and frequency of episodic spells.
2. The trigger of the dizziness perceived; set of Yes/No questions to explain the motion behaviors that set off the dizziness spells,
3. Headache and migraine-associated symptoms,
4. Tinnitus and hearing problems, and
5. History of relevant medical tests, results and procedures.

Sample screen shots of the survey are provided in Appendix I.

3.3 Medical Data and Health Information Privacy

Due to the medical nature of the survey, the study is approved by the Human Research Protections Program, the institutional review board for the Medical College of Wisconsin. As required by HIPAA privacy rules [33], personal identifiable information must be protected and is thus excluded from the data used in this research. Additional privacy efforts include the rounding of the age of the subjects to the nearest 5 years.

3.4 Survey Overview

The survey has 29 questions. As typical of vestibular assessments, there are five groups of manifestations that are common with BPPV. Each group of questions is shown in Table 3-1.

Table 3-1: Survey Question Details

Symptoms section	
Although you may experience many of the listed sensations, what is the single most noticeable part of your dizziness?	Multiple choice
Has your [dizziness] occurred once or more than once?	Once/More
With your [dizziness], have you had nausea and/or vomiting?	Y/N
During your [dizziness], have you ever had double vision?	Y/N
During your [dizziness], have you ever had blurry vision?	Y/N
Duration section	
Is your [dizziness] currently with you 24 hours a day, never stopping?	Y/N
Does your [dizziness] last seconds to 1 minute?	Y/N
Does your [dizziness] last about one hour?	Y/N
Does your [dizziness] last hours but less than 12 hours?	Y/N
Does your [dizziness] last 1 day or longer?	Y/N
Triggers section	
Is your [dizziness] typically made worse or triggered by lying down or rolling in bed?	Y/N
Is your [dizziness] typically made worse or triggered by automobile rides?	Y/N
Is your [dizziness] typically made worse or triggered by loud sounds?	Y/N
Is your [dizziness] typically made worse or triggered by sitting up or standing up?	Y/N
Is your [dizziness] typically made worse or triggered by walking on uneven ground?	Y/N
Is your [dizziness] typically made worse or triggered by supermarket aisles, malls, or tunnels?	Y/N

Table 3-2: Survey Question Details (contd.)

Triggers section (contd.)	
Is your [dizziness] typically made worse or triggered by turning your head while walking?	Y/N
Is your [dizziness] typically made worse or triggered by driving a car at night?	Y/N
Is your [dizziness] typically made worse or triggered by reaching or bending?	Y/N
Ear section	
Do you have ringing or other noise in your ears (tinnitus)?	Y/N
Do you have pain in your ears?	Y/N
Do you get frequent ear infections?	Y/N
Headache section	
Have you had a total of 5 or more bad headaches in your lifetime?	Y/N
Have you ever had a headache that throbs or pulses?	Y/N
Have you ever had nausea or vomiting with a headache?	Y/N
Have you ever had increased sensitivity to light with a headache?	Y/N
Have you ever had increased sensitivity to sounds with a headache?	Y/N
Have you ever had your [dizziness] associated with a headache?	Y/N
History section	
Have you had a hip or knee replacement?	Y/N

3.5 Survey Statistics

The survey was administered to 397 patients across four phases of the cohort study. 118 (30%) of respondents are males. The median age is 60 years old, and the average age is 61.0 years with interquartile range of 50-70 years. As shown in Table 3-1, there are 27 Yes/No questions and two multiple choice questions. These 29 questions, along with

gender and age, make up 31 attributes in the dataset. The correlation coefficients for each attribute against the BPPV diagnosis are shown in Table 3-3. The attributes with significant correlation values (at least 0.25) are shown in **bold letters**.

Table 3-3: Attributes correlation coefficient against BPPV diagnosis. Significant correlation values are in **bold letters**

Attribute	Attribute Correlation Coefficient against BPPV status	95% Confidence Interval
Symptoms		
Described as Vertigo	0.267	0.245-0.289
Once/More	0.098	0.074-0.122
Nausea/Vomiting	0.027	-0.005-0.059
Double Vision	0.085	0.055-0.115
Blurry Vision	0.150	0.112-0.188
Duration		
24/7 never stopping	0.182	0.154-0.21
Seconds to Minutes	0.242	0.206-0.278
~1 hour	0.151	0.115-0.187
Hours (<12h)	0.177	0.143-0.211
1 day or longer	0.162	0.138-0.186
Triggers for spells		
Lying down or rolling in bed	0.410	0.384-0.436
Automobile rides	0.156	0.126-0.186
Loud sounds	0.121	0.093-0.141
Sitting up/Standing up	0.02	-0.022-0.062
Walking on uneven ground	0.152	0.112-0.192
Supermarket aisles, malls or tunnels	0.267	0.237-0.297
While walking	0.142	0.104-0.180
Driving at night	0.156	0.112-0.200
Reaching/Bending	0.082	0.042-0.122

Table 3-4: (contd.) Attributes Correlation Coefficient against BPPV Diagnosis. Significant correlation values are in bold letters

Attribute	Attribute Correlation Coefficient against BPPV status	95% Confidence Interval
Hearing loss, tinnitus, and ear problems		
Ringing or other noise	0.100	0.072-0.128
Pain	0.117	0.143-0.211
Frequent Ear infections	0.046	0.012-0.080
Headache, Migraine and migraine-associated symptoms		
Had 5 bad headaches	0.196	0.170-0.222
Throbs or pulses	0.180	0.140-0.220
Nausea/Vomiting	0.146	0.114-0.178
Sensitivity to light	0.114	0.084-0.144
Sensitivity to sounds	0.148	0.104-0.192
Dizziness associated with headache?	0.105	0.069-0.141
History		
Had a hip/knee replacement?	0.087	0.047-0.127

Four attributes: “*symptoms: described as vertigo*”, “*trigger: lying down or rolling over*”, “*duration: seconds to minutes*”, and “*trigger: supermarket aisles/malls/tunnels*” are the strongest predictors. The first three are consistent with the hallmark symptoms of BPPV.

3.6 Diagnosis Statistics

Of the 397 patients to whom the survey was administered, 186 (47%) had BPPV, and the others were not identified with BPPV, labeled as “No BPPV”. Considering that the data is obtained over the space of three years, the 47% prevalence is a rough estimate of the

prevalence in the target population: those scheduled for a vestibular examination at the Medical College of Wisconsin.

3.7 Summary

The survey study that resulted in the dataset used in this work showed that patient history and symptoms information can be obtained in a structured format using questionnaires and survey tools. The Android application platform and cloud technology provided the ease of survey administration and remote access to the data. The structured data has the benefit of having less data cleaning and feature extraction work to do, as is uncommon with medical triage data. This makes machine learning and prediction easier to carry out. The prediction model results are discussed in the next chapter.

CHAPTER 4 BPPV PREDICTION RESULTS

In this chapter, prediction models derived from the logistic regression, neural networks, Naïve Bayes, and decision tree algorithms are presented and evaluated. Cross-validation output, area under the Receiver Operating Characteristic (ROC) curve, accuracy, sensitivity, and specificity are metrics used in evaluation of these prediction models.

4.1 Performance Metrics

Training is carried out using a 10-fold cross-validation process for the purpose of generalization. Performance is illustrated using the ROC curve which depicts the trade-off between sensitivity and specificity across a range of thresholds. Accuracy is defined as the fraction of correct predictions. Sensitivity is the fraction of positive samples correctly predicted, while specificity is the fraction of negative samples correctly predicted. The area under the ROC curve (AUC) value represents a summary of the ROC of a binary classifier. A better-than-chance model will have an AUC value greater than 0.5, while a perfect model will have an AUC of 1.0.

4.2 Logistic Regression

Prior to employing logistic regression to the dataset, certain preprocessing tasks were carried out. First, the symptoms question (a 7-category attribute) were transformed to binary variables, as a simplification step and consistent with the literature [28], [29]. The binary transformation result for the symptoms attribute is a categorization of either “Vertigo” or “Others”. Second, for feature extraction, attributes with statistically

insignificant correlation were excluded from the regression analysis process. Using the cross-entropy method as a cost function and a sigmoid transfer function, we obtain a logistic regression model made up of a linear predictor formula and a logistic function as in Equation 4-1.

The linear predictor formula obtained,

$$\begin{aligned}
 LP = & -1.91 + 1.61 \times (\text{Trigger: Lying down or rolling over}) \\
 & + 0.94 \times (\text{Symptoms: Vertigo}) + 0.69 \times (\text{Duration: secs to mins}) \\
 & - 0.96 \times (\text{Trigger: supermarket aisles|mall|tunnels}).
 \end{aligned}$$

4-1

The variables present in the linear predictor formula are constrained to binary values, “Yes” is designated as binary 1, while a “No” as binary 0. The linear predictor result is transformed to a probability estimate through the sigmoid function shown in Equation 4-2. For example, if the patient indicates dizziness as vertigo; triggered by rolling over in bed; spells lasting seconds to minutes; but never triggered by supermarket aisles, the formula computes as $LP = -1.91 + 1.61 + 0.94 + 0.69 - 0.96$, which equals 0.37. The probability estimate of BPPV becomes 0.59.

$$Prob(BPPV) = \frac{1}{1 + e^{-LP}}$$

4-2

We assess the logistic regression performance through a 10-fold cross validation sampling process, the accuracy, sensitivity, and specificity recorded are 0.74 ± 0.04 , 0.73 ± 0.12 and 0.75 ± 0.09 , respectively, at a decision threshold of 0.5. The Receiver

Operating Characteristics (ROC) curve is shown in Figure 4-1, with the area under the ROC curve estimated as 0.78.

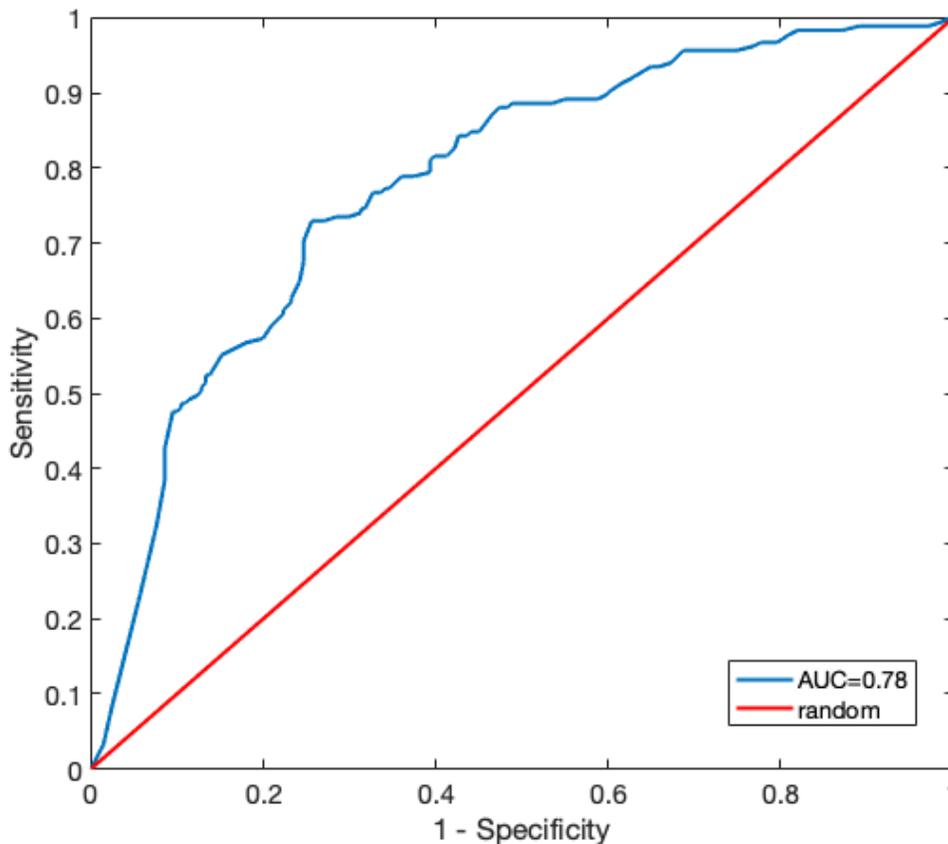


Figure 4-1: Receiver Operating Characteristics of the Logistic Regression Model, AUC=0.78

Performance accuracy of 0.74 implies that about 26% of diagnosis decisions are incorrect. This is not a satisfactory result considering that vestibular problems have significant impact on the daily routine, social, and work life of the patient. However, it is noteworthy that the weights of the variables in the regression formula correspond with their relevance in the BPPV diagnosis literature. This is an indication that underlying patterns necessary for diagnosis definitely exist in our data.

Comparing with literature, the logistic regression model obtained in [30] and [31] resulted in an AUC of 0.76, which is similar to the 0.78 obtained in this work, testifying to the similarities in both studies despite the differences in population and in the variables. It is apparent that the adjustments made in this work against the older study did not alone significantly improve model performance. For example, provisioning the questionnaire on a tablet computer did not alone show better results than the older paper format. However, computer-specific measures, such as the use of clip arts and sliding scales to express the severity of symptoms, can provide improved survey experience which, as we show in Chapter 5, may lead to better results.

4.3 Naïve Bayes algorithm

The Naïve Bayes algorithm, as discussed in Section 2.3.2, is useful in medical diagnosis [32]. To assess the influence of feature reduction with the Naïve Bayes algorithm, cross-validated training was performed on two occasions. First, on the dataset, and second, on a feature-reduced dataset. On the first occasion, we obtain an accuracy of 0.70 ± 0.10 with AUC estimate of 0.77. On the feature-reduced occasion, the selection process ruled out all but four attributes with significant class correlation (≥ 0.25) resulting in an accuracy and AUC of 0.73 ± 0.09 and 0.78, respectively, as shown in Table 4-1.

Table 4-1: Naïve Bayes cross-validated model performance with/without feature reduction

	Without Feature Selection	With Correlation Feature Selection
Accuracy	0.70 ± 0.10	0.73 ± 0.09
Sensitivity	0.69 ± 0.11	0.72 ± 0.10
Specificity	0.70 ± 0.10	0.72 ± 0.12
Area under the ROC curve	0.77	0.78

As shown in Table 4-1, the amount of discriminating information in the excluded attributes is insignificant to BPPV diagnosis. This suggests that four attributes alone can predict BPPV diagnosis successfully as all of the attributes combined. In that case, the Naïve Bayes predictions become simple likelihood formulae as follows:

Likelihood(BPPV)

$$\begin{aligned}
 &= P(BPPV) \times P(\textit{Trigger: Lying down or rolling over}|BPPV) \\
 &\times P(\textit{Symptoms: Vertigo}|BPPV) \times P(\textit{Duration: secs to mins}|BPPV) \\
 &\times P(\textit{Trigger: supermarket aisles ... |BPPV})
 \end{aligned}$$

4-3

Likelihood(No BPPV)

$$\begin{aligned}
 &= P(\textit{No BPPV}) \times P(\textit{Trigger: Lying down or rolling over}|No BPPV) \\
 &\times P(\textit{Symptoms: Vertigo}|No BPPV) \\
 &\times P(\textit{Duration: secs to mins}|No BPPV) \\
 &\times P(\textit{Trigger: supermarket aisles ... |No BPPV})
 \end{aligned}$$

4-4

To use Equations 4-3 and 4-4 for prediction, the probability entries are estimated during the training process. For example, the prevalence rate denoted by $P(BPPV)$ is 47%, and the proportion of BPPV subjects with the “Vertigo” response is 67%. In some instances, reported prevalence rates in the literature can be substituted in place of the training estimate. To predict the class to which a survey record belongs and given the computed likelihood values, the class with the higher likelihood is selected as the predicted class.

Considering that the four-attribute Bayesian model is developed with naïve assumptions of attribute-independence and equal contribution, a performance result of

0.72 accuracy is a motivation to explore more advanced and complex Bayesian models. Perhaps, attributes inter-dependences do exist, and a modelling of these relationships may provide better predictions.

4.4 Artificial Neural Network

Artificial Neural Network (ANN), as discussed in Section 2.3.3, is a popular technique for pattern recognition in medical diagnosis [11]. Prior to applying the ANN algorithm, feature reduction is carried out to reduce the dimensionality of the data, which is achieved by the use of an autoencoder.

4.4.1 Autoencoder for Dimensionality Reduction

An autoencoder produces a compressed representation of the original dataset [34]. By providing the classifier a compressed version of the dataset, the ANN classifier becomes more efficient in learning the underlying patterns in the data. First, more than half of the attributes in the dataset have insignificant correlation coefficient values and are irrelevant in the classification process. Second, there is the possibility of interdependence amongst the dataset attributes. The autoencoder is a neural network in an unsupervised learning mode. This setup is such that the original dataset can be reproduced from the compressed representation. The compression performance is evaluated by the closeness of the reproduced data to the original dataset.

As typical of ANN, non-binary categorical attributes in the data are hot encoded into binary 1's and 0's consistent with the rest of the attributes. Hot encoding is a process of converting categorical variables into a group of bits. For instance, the 7-category

attribute is coded as 7 binary outputs of 1's and 0's, and the pattern is consistent throughout the experiments. The age attribute, however, is unaltered and is the only real-valued attribute. The “Male” and “Female” genders are coded as binary 1 and 0, respectively. After hot encoding, the attribute count goes from 31 to 37.

A basic autoencoder architecture for our dataset is a three-layered artificial neural network with 37 nodes both in the input and output layers and a middle layer known as the bottleneck. A 13-node bottleneck example is shown in Figure 4-2.

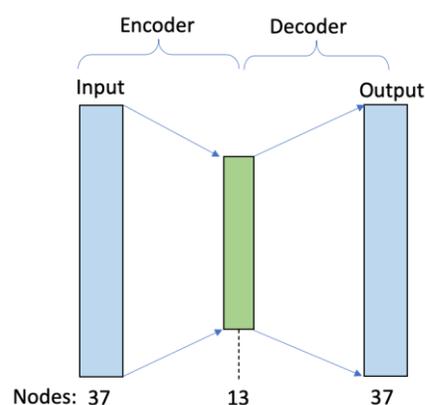


Figure 4-2: A 13-node hidden-layered autoencoder pictorial representation

To determine what autoencoder architecture is optimal in terms of reconstruction error, we experiment with a range of values for the number of nodes in the hidden layer of the autoencoder. The reconstruction error is the mean-squared-error of the attribute value difference between the reconstructed data and the original data. Given the attribute count of 37 in the dataset, the hidden layer node count is varied from 5 to 25 (roughly two-thirds of 37.) The autoencoder training process is performed with an L2 regularization parameter of 0.001 to 0.005 and sigmoid transfer functions across the layers. Table 4-2 shows the autoencoder configurations with the seven lowest

reconstruction error values. The output of the bottleneck layer serves as input to the classification stage involving a feed-forward ANN classifier.

Table 4-2: Details of the autoencoder configurations with the least reconstruction errors

Hidden layer node count	Reconstruction error
22	0.1112
24	0.1424
25	0.1611
13	0.1625
23	0.1672
17	0.1733
14	0.1743

4.4.2 ANN Classifier

In the classification stage, we train a feed-forward neural network in a 10-fold cross-validation fashion. The architecture is designed to have three layers: one hidden layer in between the input and output layers, as shown in Figure 4-3. The sigmoid transfer function is used as activation across the layers. For each output dimension of the previous autoencoder stage, we train an ANN with a hidden layer node count reasonable to those of the input and output layers. A typical range starts from a number that is the greater of (two or one-fifth of the input node count) up to the lesser of (13 or 80% of the input node count). The best performing configurations according to the area under the ROC curve, accuracy, sensitivity, and specificity values are shown in

Table 4-3.

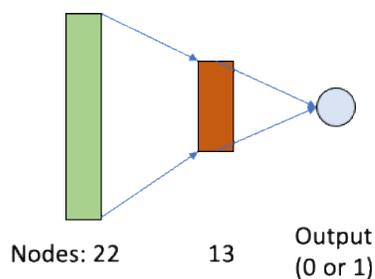


Figure 4-3: A three-layered feed-forward ANN classifier

The 22-dimension autoencoder again turned up as the best performer in terms of AUC and accuracy of 0.76 and 0.69, respectively, and nine hidden layer nodes. The corresponding sensitivity and specificity are 0.56 and 0.78, respectively.

Table 4-3: 10-fold cross validation results of various feed-forward ANN configurations on the different autoencoder outputs in the previous table (highest values for AUC and Accuracy are highlighted in red)

Feature sizes	Hidden layer nodes	AUC	Accuracy		Sensitivity		Specificity	
			Mean	SD	Mean	SD	Mean	SD
22	13	0.73	0.69	0.12	0.56	0.11	0.81	0.14
	9	0.76	0.69	0.10	0.56	0.16	0.78	0.14
	8	0.74	0.69	0.12	0.57	0.13	0.79	0.11
	6	0.72	0.67	0.12	0.55	0.16	0.78	0.14
24	7	0.72	0.67	0.10	0.53	0.11	0.78	0.10
	6	0.73	0.68	0.11	0.62	0.16	0.74	0.13
25	7	0.75	0.68	0.10	0.61	0.12	0.74	0.13
	5	0.74	0.66	0.10	0.54	0.10	0.77	0.11
13	8	0.70	0.65	0.10	0.48	0.21	0.81	0.14
	5	0.67	0.64	0.09	0.51	0.13	0.75	0.11
	3	0.68	0.62	0.11	0.50	0.11	0.78	0.14
23	7	0.73	0.69	0.10	0.57	0.13	0.84	0.11
	4	0.71	0.69	0.09	0.50	0.20	0.81	0.15

4.5 Decision Trees

Decision trees offer human readable results with the benefit of hierarchical information. Prior to generating decision trees from the data in this work, attributes with insignificant (<0.25) correlation coefficient values against the BPPV class were excluded. Because decision trees are prone to overfitting, pruning is a technique further used to sever branches that are neither significant nor improve the tree performance. The J48 algorithm [35] in the WEKA environment [36] is used to generate the decision trees in this study.

4.5.1 Correlation Attribute Evaluation

Only four attributes were found to have significant attribute correlation of at least 0.25 with the BPPV class, as seen in Table 3-3. Feature extraction involves excluding all but the four attributes, and is an important step carried out prior to applying the decision tree algorithm to the data. This reduces the possibility of overfitting and prevents or lessens the need for pruning while learning.

The decision tree output obtained on the overall dataset, 397 records, is shown in Figure 4-4. The model performance numbers are 0.69 ± 0.10 , 0.67 ± 0.13 , and 0.70 ± 0.16 for accuracy, sensitivity, and specificity, respectively. The decision path begins at the highest node of the tree and, depending on the response to the question at that node, proceeds to the next child node. A classification is realized when a leaf node is encountered, i.e., a node without a child node signifies that a decision has been reached. The decision tree in Figure 4-4 reveals the misclassification errors resulting from evaluating the tree on the dataset.

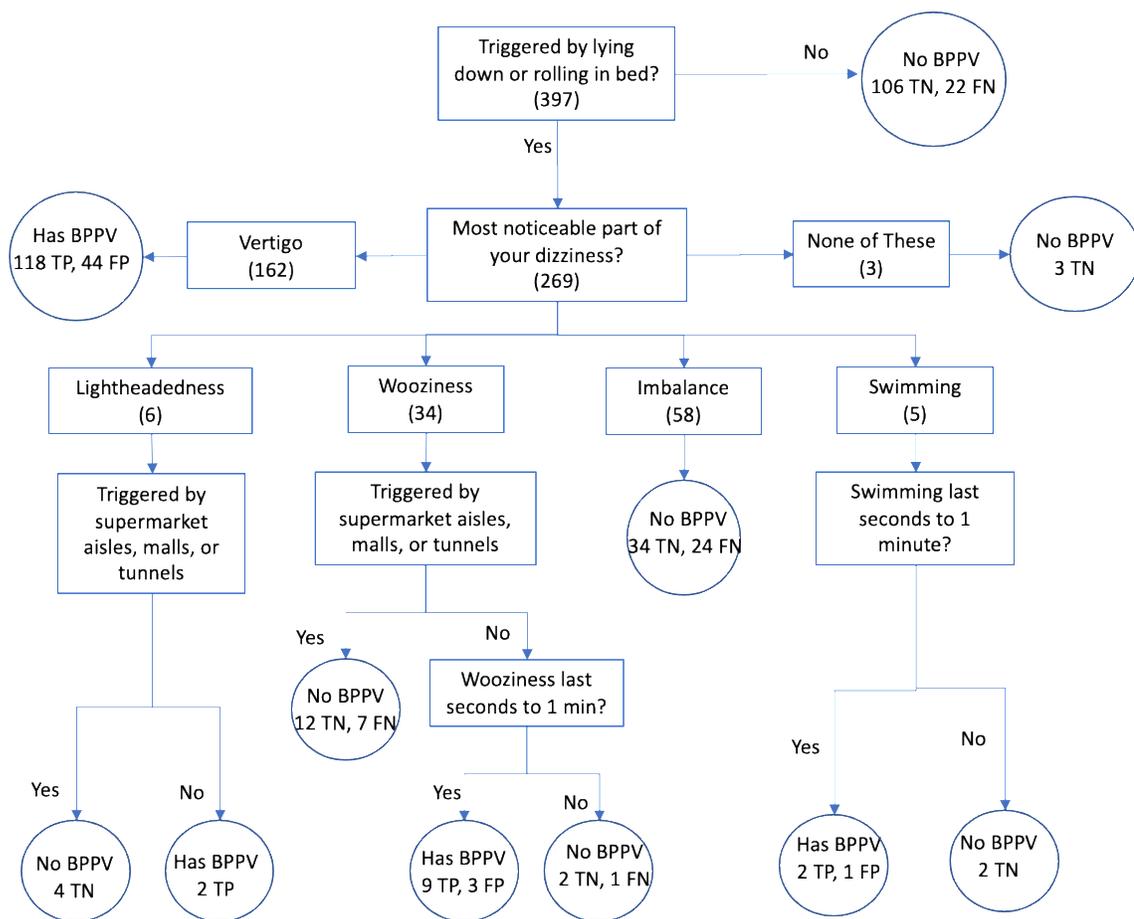


Figure 4-4: Decision tree on overall dataset

To further simplify the tree, the node with 7 branches had five child nodes collapsed into a single node. Those five nodes were combined to form a new child node named “Others”, while the remaining two branches remain, as in Figure 4-5. This is consistent with the etiology of BPPV that affirms that vertigo is a classic symptom of BPPV. The seven-category node is transformed to a binary one: “vertigo” or “others” as shown in Figure 4-5. It is worth noting that this simplification does not in any way suggest an improvement in performance because the tree simplification happens with the decision tree output and not before or during training.

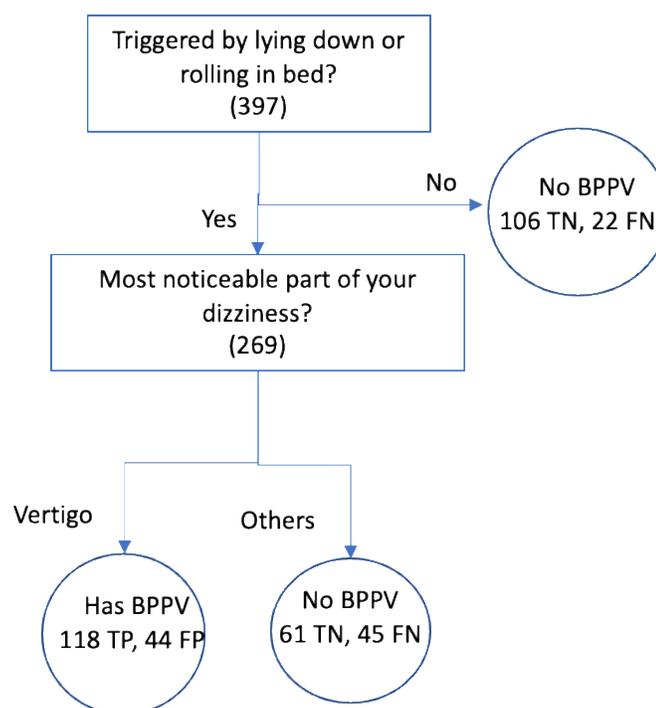


Figure 4-5: Decision tree on "Collapsed sensation" question

4.5.2 Model Performance on Single Questions

To examine the influence of each attributes to classification performance, we derive the decision tree result on each of the four prominent attributes in the dataset. Interestingly, the most successful of the four attributes alone had an accuracy of 0.65, only 0.05 less than the model learned on all the four attributes combined. This tree is shown in Figure 4-6. This single variable predictor outcome lessens the appeal of the overall tree model and could only suggest that it is possible to achieve much better tree model performance if attributes with similar predictive strengths are available in the data.

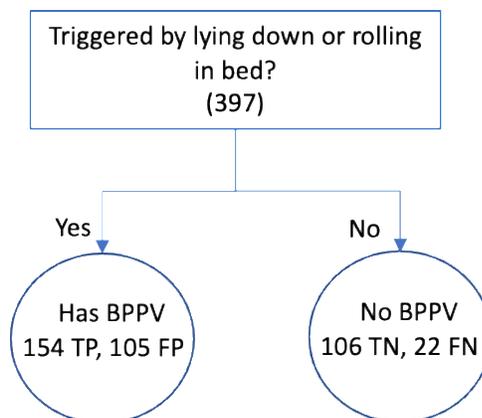


Figure 4-6: Decision tree on "Lying down/rolling over"

We find the decision tree algorithm the most appealing of the prediction algorithms discussed because of the human interpretability advantage over the others. The inherent ability to represent these trees in formats applicable to clinical systems is a desirable feature. For this reason, the decision tree model is used in further discussion later in the thesis.

4.6 Opportunities for Improvement

Despite the structural dissimilarities in the four algorithms discussed, the performance results are similar, with the accuracies 0.70 to 0.73. This discovery suggests many interesting points. First, it is possible that the limits of dataset learning have been reached. Second, it is also possible that the data is noisy, caused by either the survey approach manner or behavioral idiosyncrasies of the questionnaire respondents. Considering that the survey is meant to be taken before an appointment, patients could see the survey as a formality to obtaining an appointment rather than a useful diagnostic

tool. These insights, subsequently explored in the next chapter, can influence the survey experience and provide better prediction models.

4.7 Summary

In this chapter, four machine learning algorithms are explored on the BPPV patient-intake questionnaire data to some degree of success. The Logistic Regression model parameters showed striking consistence with expectations in the literature. The Naïve Bayes, Artificial Neural Network, and Decision Tree classifiers perform no better than the regression model, but the outcome of our research shows that there are opportunities for improvement, as discussed in the next chapter.

CHAPTER 5 IMPROVING THE USER SURVEY EXPERIENCE

In this chapter, we explore the user survey experience by examining the validity of the survey responses. We show that question misinterpretation and inconsistencies existent in the survey data is influential to prediction performance, and we propose data quality improvement measures.

5.1 Motivation

In CHAPTER 4, we evaluated the performance of four machine learning algorithms. We established that distinguishing patterns necessary for predictive diagnosis do exist and can be learned. However, the performance is disappointing, and we examine reasons for the limits of performance. We discovered certain inconsistencies in the information provided by the survey respondents and what is expected in the literature. For example, vertigo, a classic symptom of BPPV, is reported by only 67% of patients with BPPV. This is inconsistent with what is expected in the literature and reaffirms the presence of noise and errors in the survey data.

In this chapter, we explore measures for improving the quality of provided answers and show that prediction performance is influenced by this. The decision tree is used for analyses throughout this chapter for two reasons. First, prediction performance is nearly the same across the four algorithms explored, as discussed in the chapter 4. Second, the decision tree model is interpretable, and, beyond the numeric performance increase, the changes in the contents of the model are visible.

5.2 Importance of Accurate Survey Responses

It is often said in the machine learning sphere that a prediction model is only as good as its data. That is, bad or erroneous data can pose a serious limitation to predictive performance because the patterns that are recognized and learned during training can be altered by errors and noise and thereby compromise accuracy.

There are many ways to acquire patients' medical information electronically. Despite the ease of access to information, data quality is not automatic, and deliberate efforts must be made to ensure that the standards of data quality are met. A survey taker can only provide accurate information when the prompt is clearly understood. In a similar vein, medical information on symptoms and patient history is as well crucial to a successful diagnosis. Because of this, there is an emphasis in the literature on the importance of acquiring accurate patient-reported symptoms. In a vestibular study intake questionnaire [28], patients were confused on distinguishing between the individual dizziness episodes and the period of those spells, and may need follow-up questions to clarify their answers. In the same study, the manner of presentation of the question is said to play a role in the predictive ability of the data. We present insights into the survey responses and offer explanations for inconsistencies in our expectations for an accurate dataset. We show that elimination of these complications leads to better prediction models.

5.3 Average Response Time Vs. Misclassifications

We examine the average question response rate and explore relationships in the misclassification rate of the decision tree model. There are a number of reasons for

spending more time than usual in the survey. One reason is that the patient finds difficulty understanding the prompts. As in Figure 5-1, half of the respondents spent between five and eight seconds on a prompt, 15% both on 2-4 and 9-12 time slots and less than 4% on larger time durations. In Figure 5-2, the prediction accuracy within the time groups is the fraction of the ‘correct’ bar relative to the ‘Total’ bar. The decision on what time threshold to classify whether a patient understood the survey prompts may be hard to come by. As such, to use the average response time as a feature is discouraged.

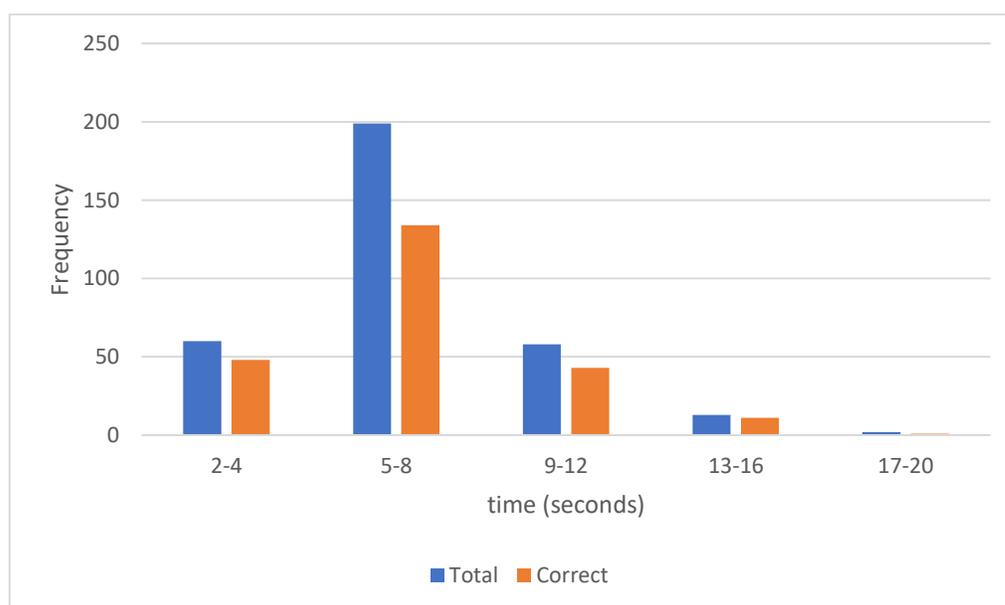


Figure 5-1: Histogram of Average Time Spent per Question.

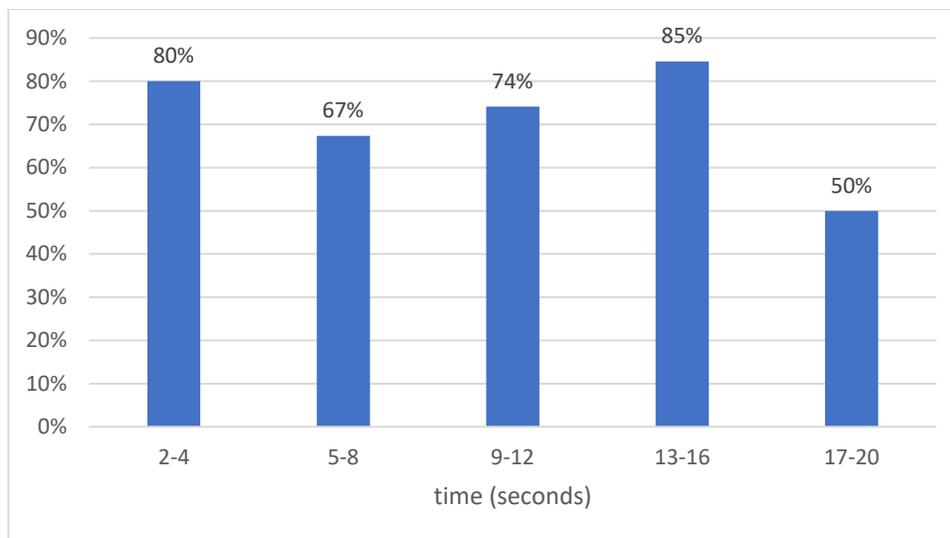


Figure 5-2: Prediction Accuracy by Average Time Spent per Question

5.4 Inaccuracies in Reporting of Dizziness Episodes Duration

As a hallmark symptom of BPPV, the length of dizziness spell is expected to be brief, lasting between 30 to 60 seconds. As such, the dizziness episodic timing is a key discriminant in diagnosing BPPV. However, in the literature, reporting accurately the length of such dizzy spells is a real challenge [20] [28] [37]. This is primarily because patients sometimes fail to distinguish the continuing discomfort after the spells from the actual dizziness episodes [29]. In some other cases, the spells might be too frequent to distinguish a series of spells from one long spell. As such, we investigate possible incidence in our dataset.

A dissection of the duration spell responses in the survey is shown in Figure 5-3. The survey had four questions regarding the duration of spells: (1) 30 seconds to 1 minute, (2) about one hour, (3) less than 12 hours, and (4) more than one day. A patient with BPPV would be expected to choose category 1, and indeed this was selected by 83%

of patients with BPPV. However, 18%, 24%, and 27% of patients with BPPV responded affirmative to categories 2, 3, and 4, respectively. In fact, 34% and 19% responded to any two and any three of the categories, respectively. These are patients who apparently did not have perfect understanding of what the prompt required at the time of the survey. Therefore, these issues contribute negatively to the prediction performance.

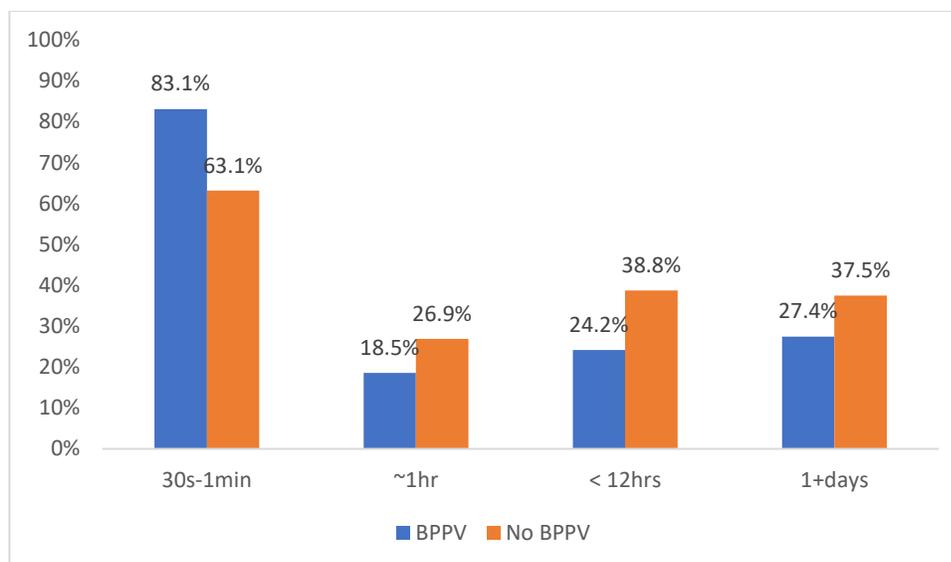


Figure 5-3: Percentage of "yes" responses to the duration questions

5.5 Symptoms Misinterpretation

Everyone has a dizzy spell now and then, but the term “dizziness” can mean different things to different people. For some, dizziness might mean a fleeting feeling of faintness, while for another, it could be an intense sensation of spinning that lasts a long time. A classic symptom of BPPV is the experience of vertigo in a brief moment. The inability to confirm or properly describe the dizziness experience is a known difficulty in BPPV diagnosis [20], [28]. The questionnaire had four options for the manifestation of dizziness: (1) Vertigo, (2) Light headedness, (3) Wooziness, and (4) Swimming sensation. A patient with BPPV would be expected to choose category 1. Indeed, this was

selected by 67% of patients with BPPV. This brings into question the integrity of the responses to this question in the survey data. Thus, it is reasonable to state that these inaccuracies contribute to the performance limitations of the models evaluated in CHAPTER 4.

5.6 Performance Improvement Measures and Simulation

In Sections 5.3 to 5.5, the integrity of the survey data was questioned. We assumed that our performance is limited by the noise and inaccuracies in the data. To test our assumptions, we handpick each of the four prominent attributes in the dataset and examine the resulting model prediction result on an adjusted version of the dataset. For instance, if we assume that BPPV patients will almost certainly report “vertigo,” then we can adjust the data to support the claim. The resulting model performance is then an estimate of the prediction performance if the survey responses were perfect.

5.6.1 The “Vertigo” Symptom Question

Let us assume that subjects diagnosed with BPPV almost certainly experience “vertigo”. To support this assumption, we adjust the dataset by setting the responses of BPPV patients to the value of “Yes.” 33% (62 of 186) of BPPV patients were affected by this adjustment process. The decision tree learned from this adjusted dataset produces accuracy, sensitivity, and specificity of 0.84, 1.00, and 0.71, respectively. This tree is shown in Figure 5-4: Tree model obtained from "Vertigo" adjusted data.

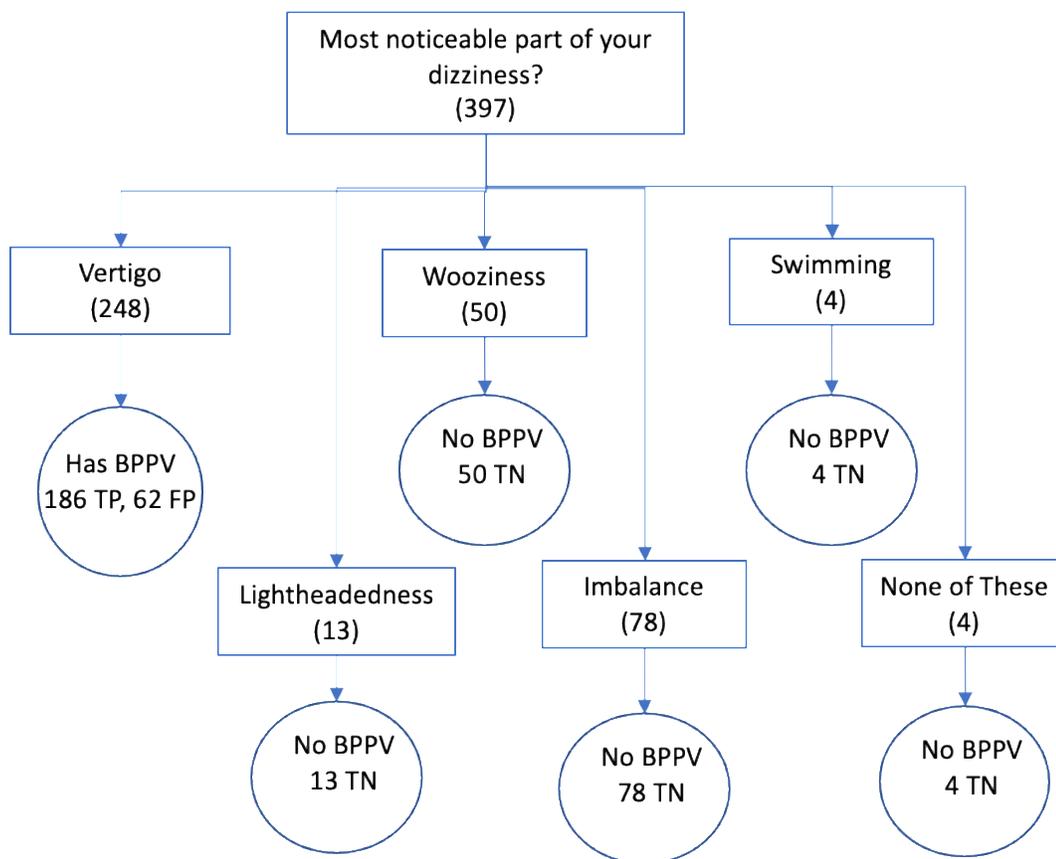


Figure 5-4: Tree model obtained from "Vertigo" adjusted data

The result of this assumption is a model with greater sensitivity than the original model, that correctly predicts BPPV as long as the BPPV patient experiences vertigo.

5.6.2 The "Lying Down/Rolling Over" Trigger prompt

In a similar way, we assume that BPPV subjects almost certainly experience dizziness triggered by lying down or rolling over in bed. Therefore, we adjust the data to reflect this assumption, by working on a copy of the original dataset. 22 (12%) of the 186 BPPV subjects had their response to this question changed from "No" to "Yes." An interesting discovery is that each of the 22 records were found to respond "Yes" to either "Sitting up/Standing up" or "Bending/Reading over". However, considering that these three

positions are very related, it is highly likely that dizziness triggered by any of the latter two would almost certainly be triggered by the lying down or rolling over position. The decision tree output from the adjusted dataset produces accuracy, sensitivity and specificity of 0.74, 0.81, and 0.68, respectively, as shown in Figure 5-5. The result of this assumption is a slightly more sensitive model than the original. A comparison with other adjustments is shown in Table 5-1.

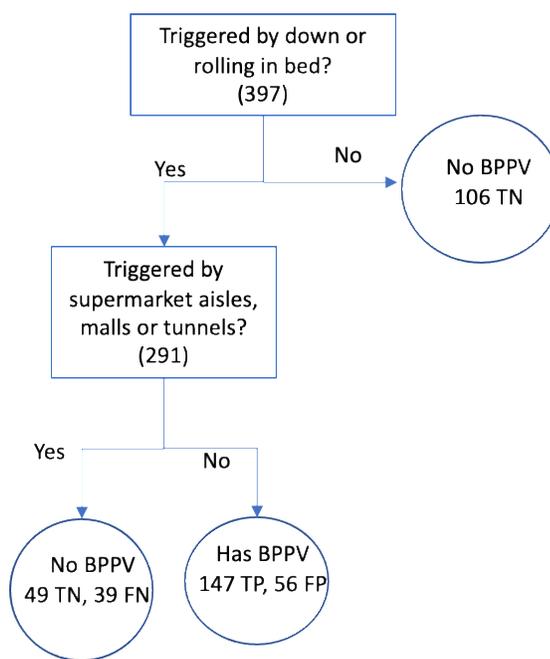


Figure 5-5: Tree model result from adjusted "Trigger: lying down" prompt

5.6.3 The Dizziness Spell Timing Prompt

We assume that BPPV patients almost certainly experience a brief episode of vertigo lasting seconds to a minute, consistent with the medical literature. Therefore, we adjust the data to reflect this assumption and examine the model performance. The adjustments were performed on a copy of the original dataset. 35 (19%) of 186 BPPV records were affected in this regard. The decision tree learned from this adjusted dataset produces an

accuracy, sensitivity and specificity of 0.76, 0.85, and 0.68 respectively. Tree model is shown in Figure 5-6. Comparison with other adjustments is shown in Table 5-1.

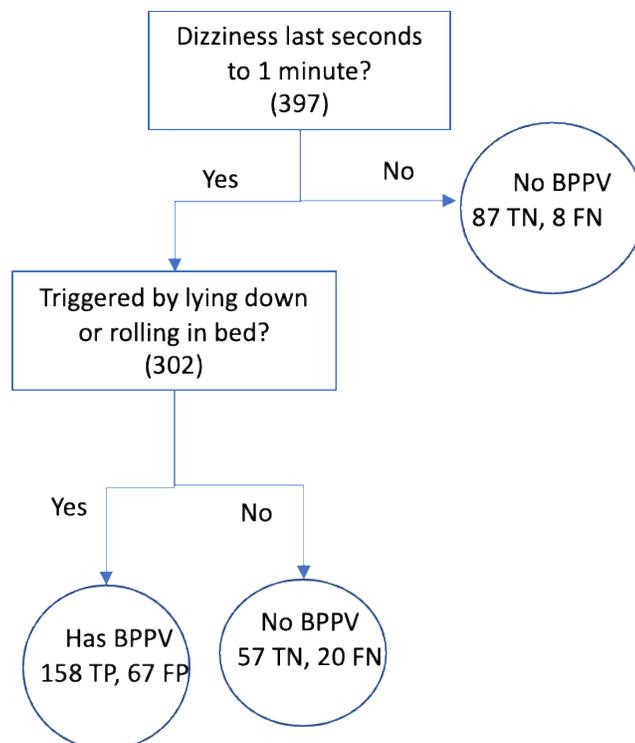


Figure 5-6: Tree model result from "Duration: brief spell" prompt

5.6.4 All Three Assumptions

We examine the outcome of our simulation if all the three adjustments discussed previously in Sections 5.6.1-5.6.3 are done simultaneously on the same set of data. This is intended to reflect the case if all the three prompts in question were responded to as we expect. The tree output produces an accuracy, sensitivity and specificity of 0.91, 0.96, and 0.87 respectively. Tree is shown in Figure 5-7. A tabular comparison of the simulation results is shown in Table 5-1.

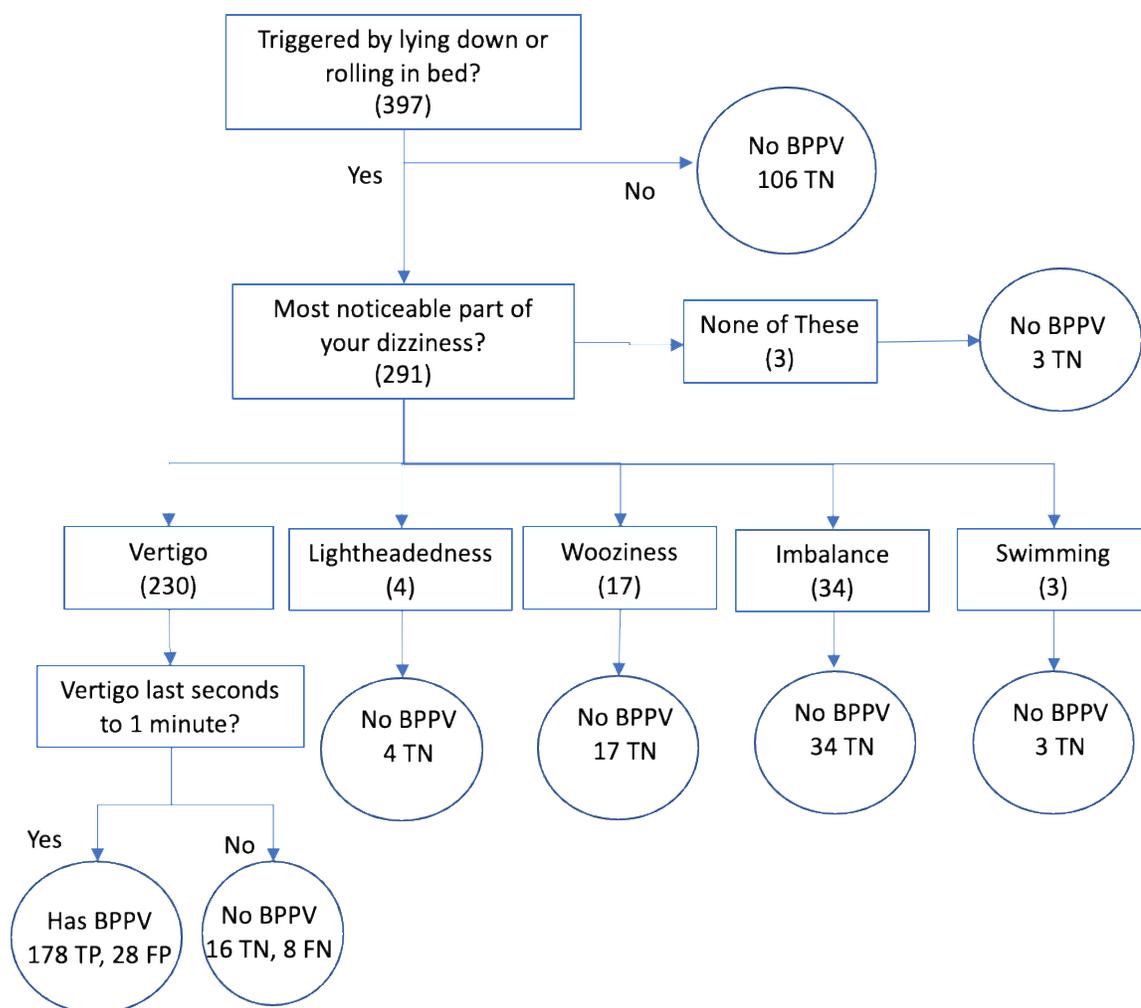


Figure 5-7: Tree model result of “All 3 attributes” adjustment

Table 5-1: Tree model performance on the adjusted dataset

Adjusted Attribute	Accuracy	Sensitivity	Specificity
No adjustments	0.69	0.67	0.70
“Vertigo” prompt	0.84	1.00	0.74
“Trigger: lying down/rolling over”	0.74	0.81	0.68
“Duration: episodic timing”	0.76	0.85	0.68
All three adjustments	0.91	0.96	0.87

Prior to this section, we questioned the integrity of the responses in the survey. By making assumptions backed by the medical literature, we simulate the performance results of the model output if the responses were accurate. The results in Table 5-1 show the individual impact of the adjustments in each attribute on the overall model performance. In terms of model accuracy, the “Vertigo” attribute showed the most improvement followed by the “Duration” and the “Trigger” attributes. Although the attributes contribute differently to the diagnosis, it is possible that the “Vertigo” prompt exhibits noise higher than the other two attributes. This is unsurprising, considering that it takes a certain level of understanding to differentiate between the different kinds of dizziness presentations namely vertigo, lightheadedness, wooziness, imbalance, and swimming sensations. The model performance of all three adjustments is a glimpse of what can be achieved if the responses to the survey are as nearly accurate as possible. Next, we present measures on alleviating these survey limitations.

5.7 Survey User Education

In the previous sections in this chapter, we established that question misinterpretation and incorrect dizziness timing estimates are responsible for inaccuracies in responses and consequently performance limits of prediction models. We, in turn, propose survey education methods that ensure that attempt to every question is well understood in terms of what is asked and what the options really mean.

To improve patient understanding of the survey, the following measures are proposed:

5.7.1 In-Survey Tutorial

BPPV is one of the numerous causes of dizziness, and since symptoms of vestibular disorders are similar and often overlap, it is important to educate the patient especially in questions that require an accurate description of the dizziness experienced. This can be accomplished by presenting one or two scenarios that depict the hallmark signs of BPPV. An example is as follows:

“My dizziness often comes when lying down or turning in bed. It feels like I am spinning or moving. The dizziness lasts less than a minute. This has happened on many days or weeks.”

5.7.2 Interface Redesign

The nature of the dizziness question that requires the patient to choose one of “Vertigo,” “Wooziness,” “Imbalance,” and “Light-headedness” is too restricting. For some patients, this is not a simple choice to make, especially if two or more options apply. To deal with this, an interface that allows the patient to provide a measure of the frequency or intensity of any of the forms of dizziness can capture this information accurately such as the use of slide bars, clip arts and smiley icons that help the patient to describe their symptoms as well as they can.

5.7.3 Language Simplification

For example, the prompt “Although you may experience many sensations, what is the single most noticeable part of your dizziness?” is a complex sentence and can be rephrased “Which of these symptoms do you experience/notice the most?”, which is

easier to understand and less formal. For medical surveys such as in this work, perhaps a sixth-grade level of reading is fitting to guarantee 100% comprehension of the prompts. To achieve this, the survey can be reviewed by a sixth-grade teacher to certify the level of comprehension.

5.7.4 Response Consistency Check

Asking for the same information multiple times but phrased differently can significantly help the patient with an accurate description of the dizziness symptoms. This suggests that follow up questions may be needed to clarify answers. However, additional questions can considerably prolong the survey time. If we make it easier for patients to describe their symptoms and find more sensitive ways to listen to them through our surveys, the data quality is surely bound to improve significantly.

5.8 Summary

In this chapter, we raised the question “how much can we trust patient’s answers?” We then present measures for improving data integrity and noise reduction in the survey data. We demonstrate that effectively reducing ambiguities in the three most important questions in the survey will result in significant improvement in model prediction performance, and by extension, BPPV diagnosis.

CHAPTER 6 CONCLUSIONS AND RECOMMENDATIONS

This chapter provides a summary of the work done in this thesis. We provide a recap of the results of the prediction models explored, along with the contributions to BPPV diagnosis research. We offer recommendations on how the ideas presented in this work can be extended, used, or applied to improve the existing performance of BPPV predictive diagnosis.

6.1 Research Contributions

In CHAPTER 4, we explored four different machine learning algorithms namely Logistic Regression, Naïve Bayes, Artificial Neural Network, and Decision Trees. We assessed them using cross-validation sampling. In the ANN model training, the autoencoder was used to compress the data into 22 attributes with a reconstruction mean squared error of 0.1112. The Correlation Feature Selection method used for feature reduction in the other algorithms was shown to eliminate all but four attributes. The classification accuracies of all four prediction models were similar, between 0.70 and 0.76. We show that the performance limits encountered are influenced by the quality of the survey responses.

6.2 Recommendations

In CHAPTER 4, we presented the prediction algorithms results and suspect that the performance is limited by the noise in the data. Further research should examine the noise robustness of the prediction models and identify those that are more tolerant to noise than the others. An ensemble strategy of two or more of the prediction models may also yield promising results. In CHAPTER 5, we proposed measures for improving the quality of

the survey responses. Intra-section tutorials and sample scenarios help the survey taker relate to their condition. Using the mobile application user interface toolbox, we can provide the patient with the ability to capture expressions of their condition. Video games and augmented reality applications are up-and-coming tools that can be used to simulate various vestibular conditions to which the patient can identify with. Language simplification is necessary for patient understanding of the prompts. Lastly, follow-up questions for clarifying and validating answers to the survey.

In CHAPTER 5, we initiated a discussion of whether response time is correlated with the uncertainty posed by the question. We, however, have no statistics for time spent on every question. Further research in this area would require measuring the time taken on every question and recording of the survey taker's clicks, footprints and progression throughout the survey.

This work is within the scope of binary classification. However, we recommend exploring multi-class prediction problems. This is because there are many vestibular disorders manifesting dizziness and whose symptoms overlap. It is possible that extending the study to a wider range of disorders will, apart from increase the target population, reveal variables with significant distinguishing power or correlation with BPPV diagnosis that might have been overlooked.

Bibliography

- [1] National Institute on Deafness and Other Communication Disorders, “Balance Disorders,” 2016. Accessed: Oct. 06, 2020. [Online]. Available: <https://www.nidcd.nih.gov/sites/default/files/Documents/health/balance/NIDCD-Balance-Disorders.pdf>.
- [2] A. M. Bronstein, J. F. Golding, M. A. Gresty, M. Mandalà, D. Nuti, A. Shetye, and Y. Silove, “The Social Impact of Dizziness in London and Siena,” *J. Neurol.*, vol. 257, no. 2, pp. 183–190, Feb. 2010, doi: 10.1007/s00415-009-5287-z.
- [3] Y. Agrawal, J. P. Carey, C. C. Della Santina, M. C. Schubert, and L. B. Minor, “Disorders of balance and vestibular function in US adults: data from the National Health and Nutrition Examination survey, 2001-2004,” *Arch. Intern. Med.*, vol. 169, no. 10, pp. 938–944, May 2009, doi: 10.1001/archinternmed.2009.66.
- [4] M. E. Tinetti and C. S. Williams, “Falls, injuries due to falls, and the risk of admission to a nursing home,” *N. Engl. J. Med.*, vol. 337, no. 18, pp. 1279–1284, Oct. 1997, doi: 10.1056/NEJM199710303371806.
- [5] D. Fife and J. E. Fitzgerald, “Do patients with Benign Paroxysmal Positional Vertigo receive prompt treatment? Analysis of waiting times and human and financial costs associated with current practice,” *Int. J. Audiol.*, vol. 44, no. 1, pp. 50–57, Jan. 2005, doi: 10.1080/14992020400022629.
- [6] Gorina T, Hoyert D, Lentzner H, and Gounding M, “Trends in causes of death among older persons in the United States (October 2005).” Accessed: Oct. 15, 2020. [Online]. Available: <http://unstats.un.org/unsd/demographic/>.
- [7] G. Royl, C. J. Ploner, and C. Leithner, “Dizziness in the emergency room: diagnoses and misdiagnoses,” *Eur. Neurol.*, vol. 66, no. 5, pp. 256–263, Nov. 2011, doi: 10.1159/000331046.
- [8] T. P. Exarchos, G. Rigas, A. Bibas, D. Kikidis, C. Nikitas, F. L. Wuyts, B. Ihtijarevic, L. Maes, M. Cenciarini, C. Maurer, N. Macdonald, D. E. Bamiou, L. Luxon, M. Prasinou, G. Spanoudakis, D. D. Koutsouris, and D. I. Fotiadis, “Mining balance disorders’ data for the development of diagnostic decision support systems,” *Comput. Biol. Med.*, vol. 77, pp. 240–248, Oct. 2016, doi:

- 10.1016/j.compbimed.2016.08.016.
- [9] E. Kentala and S. D. Rauch, “A practical assessment algorithm for diagnosis of dizziness,” *Otolaryngol. - Head Neck Surg.*, vol. 128, no. 1, pp. 54–59, Jan. 2003, doi: 10.1067/mhn.2003.47.
- [10] K. Viikki, E. Kentala, M. Juhola, and I. Pyykkö, “Decision tree induction in the diagnosis of otoneurological diseases,” *Med. Inform. Internet Med.*, vol. 24, no. 4, pp. 277–289, 1999, doi: 10.1080/146392399298302.
- [11] E. Kentala, I. Pyykkö, Y. Auramo, and M. Juhola, “Neural networks in neurotologic expert systems,” in *Acta Oto-Laryngologica, Supplement*, 1997, no. 529, pp. 127–129, doi: 10.3109/00016489709124102.
- [12] K. Viikki, E. Kentala, M. Juhola, I. Pyykkö, and P. Honkavaara, “Generating decision trees from otoneurological data with a variable grouping method,” *J. Med. Syst.*, vol. 26, no. 5, pp. 415–425, Oct. 2002, doi: 10.1023/A:1016463032661.
- [13] Vestibular Disorders Association, “Vestibular disorders: an overview.” Accessed: Oct. 10, 2020. [Online]. Available: <https://vestibular.org/article/what-is-vestibular/about-vestibular-disorders/>.
- [14] R. Domingues and M. Waseem, “Dix-Hallpike Maneuver,” in *Atlas of Emergency Medicine Procedures*, Springer, New York, 2016, pp. 257–258.
- [15] V. A. Stanton, Y. H. Hsieh, C. A. Camargo, J. A. Edlow, P. Lovett, J. N. Goldstein, S. Abbuhl, M. Lin, A. Chanmugam, R. E. Rothman, and D. E. Newman-Toker, “Overreliance on symptom quality in diagnosing dizziness: Results of a multicenter survey of emergency physicians,” *Mayo Clin. Proc.*, vol. 82, no. 11, pp. 1319–1328, 2007, doi: 10.4065/82.11.1319.
- [16] T. Brandt, D. Huppert, J. Hecht, C. Karch, and M. Strupp, “Benign Paroxysmal Positioning Vertigo: A Long-term follow-up (6-17 years) of 125 patients,” *Acta Otolaryngol.*, vol. 126, no. 2, pp. 160–163, Feb. 2006, doi: 10.1080/00016480500280140.
- [17] J. A. Lopez-Escamez, M. J. Gamiz, A. Fernandez-Perez, M. Gomez-Fiñana, and I. Sanchez-Canet, “Impact of treatment on health-related quality of life in patients with posterior canal benign paroxysmal positional vertigo,” *Otol. Neurotol.*, vol. 24, no. 4, pp. 637–641, 2003, doi: 10.1097/00129492-200307000-00018.

- [18] Mayo Clinic Staff, “Canalith repositioning procedure.” Accessed: Jul. 17, 2020. [Online]. Available: <https://www.mayoclinic.org/tests-procedures/canalith-repositioning-procedure/about/pac-20393315>.
- [19] T. Brandt, “Vertigo, dizziness, and falls in the elderly,” in *Vertigo*, Springer New York, 2003, pp. 385–392.
- [20] D. E. Newman-Toker, “Symptoms and signs of neuro-otologic disorders,” *CONTINUUM Lifelong Learning in Neurology*, vol. 18, no. 5. Continuum (Minneapolis, Minn), pp. 1016–1040, Oct. 2012, doi: 10.1212/01.CON.0000421618.33654.8a.
- [21] K. A. Kerber and D. E. Newman-Toker, “Misdiagnosing dizzy patients: common pitfalls in clinical practice,” *Neurologic Clinics*, vol. 33, no. 3. W.B. Saunders, pp. 565–575, Aug. 01, 2015, doi: 10.1016/j.ncl.2015.04.009.
- [22] “Benign Paroxysmal Positional Vertigo (BPPV) - diagnosis and treatment - Mayo Clinic.” Accessed: Sep. 25, 2020. [Online]. Available: <https://www.mayoclinic.org/diseases-conditions/vertigo/diagnosis-treatment/drc-20370060>.
- [23] K. A. Kerber, “Vertigo and dizziness in the emergency department,” *Emerg. Med. Clin. North Am.*, vol. 27, no. 1, pp. 39–50, Feb. 2009, doi: 10.1016/j.emc.2008.09.002.
- [24] E. Mira, A. Buizza, G. Magenes, M. Manfrin, and R. Schmid, “Expert systems as a diagnostic aid in otoneurology,” *ORL*, vol. 52, no. 2, pp. 96–103, 1990, doi: 10.1159/000276113.
- [25] C. Gavilán, J. Gallego, and J. Gavilán, “‘CARRUSEL’: An expert system for vestibular diagnosis,” *Acta Otolaryngol.*, vol. 110, no. 3–4, pp. 161–167, 1990, doi: 10.3109/00016489009122532.
- [26] K. Higashi-Shingai, T. Imai, T. Kitahara, A. Uno, Y. Ohta, A. Horii, S. Nishiike, T. Kawashima, T. Hasegawa, and H. Inohara, “Diagnosis of the subtype and affected ear of Benign Paroxysmal Positional Vertigo using a questionnaire,” *Acta Otolaryngol.*, vol. 131, no. 12, pp. 1264–1269, Dec. 2011, doi: 10.3109/00016489.2011.611535.
- [27] H. J. Kim, J. M. Song, L. Zhong, X. Yang, and J. S. Kim, “Questionnaire-based

- diagnosis of Benign Paroxysmal Positional Vertigo,” *Neurology*, vol. 94, no. 9, pp. e942–e949, Mar. 2020, doi: 10.1212/WNL.00000000000008876.
- [28] D. R. 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. 351–356, Apr. 2016, doi: 10.1001/jamaoto.2015.3663.
- [29] C. J. Britt, B. K. Ward, Y. Owusu, D. Friedland, J. O. Russell, and H. M. Weinreich, “Assessment of a statistical algorithm for the prediction of Benign Paroxysmal Positional Vertigo,” *JAMA Otolaryngol. - Head Neck Surg.*, vol. 144, no. 10, pp. 883–886, Oct. 2018, doi: 10.1001/jamaoto.2018.1657.
- [30] H. Richburg, R. Povinelli, and D. Friedland, “Direct-to-patient survey for diagnosis of Benign Paroxysmal Positional Vertigo,” in *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, Jan. 2019, pp. 332–337, doi: 10.1109/ICMLA.2018.00056.
- [31] J. I. E. Hoffman, “Chapter 33 - Logistic regression,” in *Biostatistics for Medical and Biomedical Practitioners*, Elsevier, 2019, pp. 581–589.
- [32] Y. Shen, Y. Li, H. T. Zheng, B. Tang, and M. Yang, “Enhancing ontology-driven diagnostic reasoning with a symptom-dependency-aware Naïve Bayes classifier,” *BMC Bioinformatics*, vol. 20, no. 1, p. 330, Jun. 2019, doi: 10.1186/s12859-019-2924-0.
- [33] HHS Office of Civil RRS, “HIPAA administrative simplification regulation text.” Accessed: Oct. 06, 2020. [Online]. Available: <https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/administrative/combined/hipaa-simplification-201303.pdf>.
- [34] J. Schmidhuber, “Deep Learning in neural networks: an overview,” *Neural Networks*, vol. 61, pp. 85–117, Apr. 2014, doi: 10.1016/j.neunet.2014.09.003.
- [35] S. L. Salzberg, “C4.5: Programs for machine learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993,” *Mach. Learn.*, vol. 16, no. 3, pp. 235–240, Sep. 1994, doi: 10.1007/bf00993309.
- [36] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data mining: Practical machine learning tools and techniques*, 3rd ed. Morgan Kaufmann, San Francisco (CA),

2016.

- [37] R. Lapenna, M. Faralli, M. R. Del Zompo, L. Cipriani, P. D. Mobaraki, and G. Ricci, “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. 881–888, Oct. 2016, doi: 10.1007/s40520-015-0487-0.

APPENDIX I Selected Screenshots from the BPPV Survey

