CHAPTER 1 INTRODUCTION

* 1. INTRODUCTION

The amount of data in the world, in our lives seems to be increasing and there is no end in sight omnipresent. As the volume of data increases, the proportion of it that people “understand” decreases alarmingly. Lying hidden in all this data is information, potentially useful information, which is rarely made explicit or taken advantage of people have been seeking patterns in the data. As the flood of data swells and machines that can undertake the searching becomes commonplace, the opportunities for data mining increases. As the world grows in complexity, data mining becomes our only hope for elucidating the patterns that underlie it. Intelligently analyzed data is a valuable resource. It can lead to new insights in marketing, surveillance, fraud detection, scientific discovery and in diagnosis of diseases, commercial settings to competitive advantages.

*Data mining* or *knowledge discovery* refers to the non trivial process of finding interesting, previously unknown and potentially useful patterns in large repositories of data. The term *data mining* also refers to the step in the knowledge discovery process in which special algorithms are employed in hopes of identifying *interesting* patterns in the data. These interesting patterns are then analyzed yielding *knowledge.* The desired outcome of data mining activities is to discover knowledge that is not explicit in the data and to put that knowledge to use[1].

Data mining refers to the process of finding interesting patterns in data that are not explicitly part of the data[2]. The interesting patterns can be used to tell us something new and to make predictions. The process of data mining is composed of several steps including selecting data to analyze, preparing the data, applying the data mining algorithms, and then interpreting and evaluating the results. Sometimes the term *data mining* refers to the step in which the data mining algorithms are applied. This has created a fair amount of confusion in the literature. But more often the term is used to refer the entire process of finding and using interesting patterns in data[3]. Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets[4]. These tools can include statistical models, mathematical algorithms, and machine learning methods (algorithms that improve their performance automatically through experience, such as Neural networks or decision trees).

* 1. 1.2 DATA MINING IN MEDICINE

Modern medicine generates a great deal of information stored in the medical database. For example, medical data may contain MRIs, signals like ECG, clinical information like blood sugar, blood pressure, cholesterol levels, etc., as well as the physician's interpretation. Extracting useful knowledge and providing scientific decision-making for the diagnosis and treatment of disease from the database increasingly becomes necessary. Data mining in medicine can deal with this problem. It can also improve the management level of hospital information and promote the development of telemedicine and community medicine[5]. The goal of data mining in clinical medicine is to derive models that can use patient specific information to predict the outcome of interest and to thereby support clinical decision-making.

Data mining methods may be applied to the construction of decision models for procedures such as prognosis, diagnosis and treatment planning, which once evaluated and verified may be embedded within clinical information systems.

The general principle involved in medical data mining is as shown in Figure 1.1. It consists of collection of patient’s records, from various sources which include patients and care givers interviews and this is the starting point in the learning system, as quality of raw data determines the overall performance. The aim of preprocessing data is to remove the noise, outliers and to discover the important features existing in the raw data. Pre-processing stage includes cleaning, normalization, transformation, and feature selection. Learning becomes simpler if features are identified at pre-processing stage. The product of data pre-processing is the training set. With the training set, the learning model has to learn from it. Once the model is developed the next stage would be validating the model using test dataset.

* 1. 1.3 CLASSIFICATIONS AS A LEARNING TASK

Medical decision making can be seen as classification: a physician classifies the symptoms of a patients to a certain disease group on the basis of the knowledge. Thus learning classification models for the various neurodegenerative disorders is the learning task in this study.

General classification models can be learned by analyzing data. Training data consists of cases (examples, instances, or objects) that are described using vectors of attributes (variables, or features). Attributes may be qualitative (measured with a nominal or ordinal scale) or quantitative (measure with an interval or a ratio scale)[6]. In supervised learning, cases usually belong to one of the mutually exclusive cases, and the class information is utilized in learning. Attributes are adequate for the classification task, if all the cases having identical attribute vectors belong the same class[7].

The classification data model constructed is, on the other hand, a description of the training data, and on the other hand a classification rule that can be applied to new cases if the training data form a representative sample from the object space.

* 1. 1.4 MACHINE LEARNING

The ability to learn is one of the most essential characteristics of intelligent behavior. Machine learning is a subfield of AI studying computational methods that can improve performance on some task by learning[8]. The aim of machine learning research may be cognitive, technical or theoretical[9]. Cognitive aims seek to model human learning at some level. Automating the process of knowledge acquisition for knowledge based systems is an example of a technical aim. Theoretical analysis considers, for example, characteristics of learning methods such as their scope and limitations. Like AI, machine learning is an inherently interdisciplinary filed. Statistic, for example, is widely utilized in the field of ML.

Machine learning methods can be classified on the basis of various criteria such as underlying learning strategy, representation of knowledge, or application domain. Langley and Simon[10] found five major paradigms in the field of Machine learning: Neural networks, Instance based learning, Genetic algorithms, Inductive learning and analytical learning. These paradigms share the common goal of improving the performance of some task, which is typically achieved by finding and exploiting regularities in training data.

Knowledge acquisition and data mining are important application areas of Machine learning. Machine learning methods have been utilized in a wide variety of application domains such as credit card fraud detection, handwritten character recognition, speech recognition, marketing analysis, quality control in manufacturing, airline seating allocation, food and chemical formula optimization and automatic classification of celestial objects[11].



Pre-processing of Records (Data Cleaning, Feature Selection)

Training Set

Applying Machine Learning Methods

Collection of Patients Medical Records

Computer Model

Figure 1.1: Medical Mining Process

The medical domains in which ML has been used are diagnosis of acute appendicitis[12], diagnosis of dermatological disease[13], diagnosis of female urinary incontinence[14], diagnosis of thyroid diseases[15], finding genes in DNA[16], outcome prediction of patients with severe head injury[17], prediction of metabolic and respiratory acidosis in children[18], as well as relating clinical and neurophysiologic assessment of spasticity[19], among many others.

* 1. 1.5 PROCESS OF APPLYING MACHINE LEARNING

Langley and Simon identify the following main phases in the process of applying machine learning methods: formulating the problem into a form suitable for ML methods selected, determining the representation, collecting the training data, evaluating the learned knowledge, and fielding the learned knowledge (i.e., taking into use).

* PRE-PROCESSING

Training data forms the basis for the learning process. The learning method can find only the concepts included in the training data, and thus, selected attributes and instances should cover different situations appearing in the problem domain. Constructing the training data is an essential part of applying ML techniques and it assumes the involvement of both a knowledge engineer and a domain expert.

* FEATURE SUBSET SELECTION

Determining the representation of learning data, that is the attributes used to describe training cases, is an important task. Although a larger attribute set usually carries more information than a smaller one, it does not necessarily produce better learning results. Increasing the number of attributes probably increases the amount of noise and missing data. Additional, possibly redundant or irrelevant attributes may also interfere with the generalization capability of ML method[20]. Therefore, a carefully selected sub set of attributes may produce better results. Generated models may be simpler, easier to understand, and more accurate. However, finding of a good subset may be a difficult and tedious task due to the wealth of attributes available, which calls for automated methods.

Various methods have been developed for Feature Subset Selection (FSS)[21]. These methods follow embedded, filter, and wrapper approaches for selecting attributes[22]. In the first approach, the selection of attributes is embedded in the learning algorithm, itself. Examples of such algorithms are ID3[23], C4.5[24], and CART[25]. In the filter approach, useful attributes are filtered before the actual learning, and the filtering is based on the characteristics of the training data. The simplest way is to choose attributes having the strongest associations with the class attribute. Another way is to find a minimal set of attributes that are adequate for the discrimination of classes.

In addition to the above three explicit FSS approaches, there exist implicit weighting methods for defining the usefulness of attributes. Tuning the weights of Neural networks with algorithms such as Backpropagation as well as memory based reasoning, are examples of this.

* DATA TRANSFORMATION

Many ML algorithms make assumptions about the scales of attributes. ID3, for example, requires categorical attributes and binary class attributes, where as a nearest neighbor classifier with the Euclidean distance function works best with quantitative attributes[26]. Sometimes, the ordinal attributes can be used like quantitative ones by assigning consecutive numerical scores to the ordinal categories[27], and nominal attributes can be transformed into binary dummy attributes. If learning requires a discrete feature base, quantitative attributes can be coded as qualitative ones. A frequently used and perhaps the simplest method for discretisation is equal width interval Binning[28]. Discretisation results in loss of information, but this is not necessarily harmful for learning results in an empirical comparison of discretisation methods, discrimination prior to learning significant improved the accuracy of induction methods in some tasks.

* SELECTION OF TRAINING CASES

In its extremes, training data may be a ‘raw’ sample extracted from a large real world dataset or a set of instances carefully chosen by an expert. If the sample is large enough, it gives a statistically reliable picture of the task to be learned, but may not be optimal for learning purposes. Data may contain redundant cases and rare cases essential for the problem solving may not be presented. In training data selected by an expert, every instance has a special meaning for the concept to be learned.

Behind the current theory on ML, there is an assumption that distributions of training and test data are similar. Good results on some distribution do not necessarily guarantee good results on other distributions and, thus, learning is not reliable when the distribution of the training data is as close as possible to the real world distribution.

In real world learning tasks, data are typically noisy, that is, attribute values and class labels may be erroneous. Noisy data result in overly complex, incomprehensible models and possibly decreased accuracy. To remove the effects of noise, methods for simplifying learned models can be used. An alternative approach is to discard the noisy cases from the training data. In general, noisy cases are in a way exceptional compared to other data.

* EVALUATION

An objective way (with respect to the chosen criteria) for evaluating learning output is the use of performance measures. The most common measure used to characterize the performance of a classifier is accuracy[29]. Accuracy(*ACC*) is calculated as the percentage of correctly classified cases:

∑ *tposc*

*ACC* = 100 % Where *c* = 1 to *C*

∑ *posc*

Where *C* is the number of classes, *tposc* is the number of correctly classified cases in the class *c*, *posc* is the total number of cases in *c*, and the error rate measures the proportion of misclassified cases.

If the class distribution is highly imbalanced, the accuracy does not give the whole picture of the classifiers quality. The accuracy may be high, even though the cases of small classes are poorly identified[30]. Therefore the true positive rate if calculated separately for each class in order to obtain the more detailed information about the classifier’s performance.

The true positive rate for class *c* (*TPRC*) is calculated as the percentage of correctly classified cases of the class:

*tposc*

*TPRC* = 100 %

*posc*

In the case of a multi-valued class label, the classifiers may also be constructed in the form of one class(the positive class) verses the other classes (the negative class). Then, the identification of the negative cases is measured by the true negative rate.

The data are typically divided into training and test sets. The classifier is constructed from the training set and tested with the separate test set. Descriptive performance measures are calculated from the training set and predictive performance measures from the test set. Due to the small amount of data available, a technique called *N*-fold cross-validation can be used instead of a separate test set to estimate the predictive performance. The data are divided into *N* subsets with a nearly equal size and a class distribution of the original data. The classifier is constructed from *N* – 1 subsets, and the remaining subset is used as the test set. This is repeated *N* times such that each subset is once the test set, and accordingly, predictive performance measures are calculated as the average of *N* experiments. Cross-validation can be repeated several times for random partitions. The average measures calculated from the results of different cross-validation times are fairly reliable estimates for the performance of a classifier constructed from the whole training data.

Even if experts are not capable of fully articulating their knowledge they can evaluate the quality and correctness of the learned models. The models should be intelligible and the reasonable from the viewpoint of the experts in order to be called knowledge.

* 1. MACHINE LEARNING IN MEDICAL ANALYSIS

As medical information systems in modern hospitals and medical institutions become larger and larger, it causes great difficulties in extracting useful information for decision support, especially when traditional manual data analysis has become inefficient and methods for efficient computer-based analysis are indispensable. Therefore, introduction of a modern, efficient and effective computer-based method in medical analysis for decision support is warranted. Medical analysis using machine learning techniques has been implemented for the last two decades. It has been proven that the benefits of introducing machine learning into medical analysis are to increase diagnostic accuracy, to reduce costs and to reduce human resources. For example, “Dementia due to Alzheimer’s disease and other dementias constitute the fourth most common disorder among the elderly and have a total cost in the USA of $100 billion annually. Proper treatment can reduce this cost by up to 25%[31]. Machine learning involves using computers to find rules, or computer models that best describe problems from an electronic database. Machine learning techniques are divided into two categories: supervised learning and unsupervised learning. As the name may suggest, supervised learning is a machine learning technique that finds the best described computer model from a database with the correct class variable. The class variable is the variable of a database that the computer model needs to classify. In the present study we used different kinds of machine learning techniques: Neural networks[32], Decision trees[33], Bayesian networks[34] and Genetic algorithms[35] etc.

Moreover, there are two different phases in any machine learning technique: training phase and testing phase. The training phase is when a computer model is induced from a set of samples in the database. In the testing phase the built computer model is tested from a set of unseen samples in the database (Figure 1.2).



Training Phase

Database

Machine Learning Algorithm

Computer Model

Testing Phase

Unseen new data

Computer Model

Classification performance

Figure 1.2: Machine Learning System

CHAPTER 2 LITERATURE SURVEY

* 1. 2.1 OVERVIEW

In this chapter, a review of the current literature on Alzheimer’s disease, applications of Data mining techniques in healthcare, a brief survey on classification of cognitive states, use of rough sets theory in medical domain, various neurodegenerative disorders, screening and diagnosis of Alzheimer’s Disease (AD) using computer tomography of brain and methods of assessment of various stages of AD using neuropsychological tests have been discussed.

* 1. 2.2 DATA MINING TECHNIQUES IN HEALTHCARE

Data mining, *the extraction of hidden predictive information from large databases*, is a powerful new technology with great potential in health care. Data mining tools predict stages of the progression of diseases and helps in taking the clinical decisions. Data mining in medicine is an emerging field of high importance for providing prognosis and a deeper understanding of the classification of various diseases. Data mining applications offers a principled approach for developing sophisticated, automatic, and objective algorithms for analysis of high-dimensional and multimodal biomedical data[47]. There have been several advances in the state of the art that have shown promise in improving detection, diagnosis, and therapeutic monitoring of diseases. The recent developments in Machine learning, focusing on supervised and unsupervised linear methods, have made significant impacts in the detection and diagnosis of diseases in biomedicine.

Recently, several Machine learning techniques such as Decision tree induction, Rough Sets (RS), Soft computing techniques, Neural networks, Genetic algorithms etc. are gaining popularity for predictive modeling. The efforts are on to apply Machine Learning to biology. In 2002, a research company correlogic stunned the medical world with the announcement of a vastly improved test for detecting ovarian cancer. Techniques in image similarity can be used to improve the classification of breast cancer images. Image feature clustering were done to reduce the noise and the feature space, and the results were used in a distance function that uses a learned threshold in order to produce a classification of Breast cancer images in the mammogram[48].

Mu-Jung Huang et al., suggested a model of chronic diseases prognosis and diagnosis system by integrating Data mining and Case-Based Reasoning (CBR)[49]. A proficient methodology for the extraction of significant patterns from the heart disease warehouses for heart attack prediction using Data mining and Artificial Neural Network is presented in[50]. Latha Parthiban et al., projected an approach on basis of Co-Active Neuro-Fuzzy Inference System (CANFIS) for prediction of heart disease. The CANFIS model diagnosed the presence of disease by merging the neural network adaptive capabilities and the Fuzzy logic qualitative approach and further integrating with Genetic Algorithm. On the basis of the training performances and classification accuracies, the performances of the CANFIS model were evaluated[51].

Ischaemic heart disease is one of the world's most important causes of mortality**.** An application of Machine learning in the diagnosis of ischaemic heart disease proved efficient[52]. Data mining methods may aid the clinicians in the predication of the survival of patients and in the adaptation of the practices consequently. The work of Franck Le Duff et al.,[53] might be executed for each medical procedure or medical problem and it would be feasible to build a decision tree rapidly with the data of a service or a physician. Comparison of traditional analysis and Data mining analysis illustrated the contribution of the Data mining method in the sorting of variables and concluded the significance or the effect of the data and variables on the condition of the study.

Genome-wide transcription profiling is a powerful technique for studying the enormous complexity of cellular states. The data obtained from high-density microarrays is highly complex and poses considerable challenges in data mining.

P. Roy Walker[54] applied various Data mining techniques for dealing with problem of microarray data that come from two known classes (Alzheimer and normal). They had applied three separate techniques to discover genes associated with Alzheimer disease.

* 1. 2.3 MACHINE LEARNING METHODS

In this section, a detail review of the most common Decision tree algorithms implemented serially, experimentations to compare their classification and prediction accuracy have been discussed.

* + 1. 2.3.1 CLASSIFICATION

Classification can be described as a supervised learning algorithm in the Machine learning process. It assigns class labels to data objects based on prior knowledge of class which the data records belong. It is a Data mining technique, has made it possible to co-design and co-develop software and hardware, and hence, such components[55]. However, integration deals with knowledge extraction from database records and prediction of class label from unknown data set of records. In classification a given set of data records is divided into training and test data sets. The training data set is used in building the classification model, while the test data record is used in validating the model. The model is used to classify and predict new set of data records that is different from both the training and test dataset[56-57]. Supervised learning algorithm (like classification) is preferred to unsupervised learning algorithm (like clustering) because its prior knowledge of the class labels of data records makes feature/attribute selection easy and this leads to good prediction/classification accuracy.

Some of the common classification algorithms used in Data mining and decision support systems are: Neural networks[58], Logistic regression[59], Decision tree[60] etc. Among these classification algorithms Decision tree algorithms is the most commonly used because it is easy to understand and cheap to implement. It provides a modeling technique that is easy for human to comprehend and simplifies the classification process[61]. Most Decision tree algorithms can be implemented in both serial and parallel form while others can only be implemented in either serial or parallel form. Parallel implementation of decision tree algorithms is desirable in- order to ensure fast generation of results especially with the classification/prediction of large data sets, it also exploits the underlying computer architecture[62]. But serial implementation of decision algorithm is easy to implement and desirable when small-medium data sets are involved.

* + 1. 2.3.2 DECISION TREE ALGORITHM

Decision tree induction represents a simple and powerful method of classification which generates a tree and a set of rules, representing the model of different classes, from a given dataset[63]. Decision tree is a flow chart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test and each leaf node represents the class. The top most node in a tree is the root node. For DT induction, ID3 algorithm and its successor C4.5 algorithm by Quinlan(1993) are widely used. Decision tree algorithm is a Data mining induction technique that recursively partitions a data set of records using depth-first greedy approach or breadth-first approach until all the data items belong to a particular class. A Decision tree structure comprises of root, internal and leaf nodes. The tree structure is extensively used in classifying unknown data records. At each internal node of the tree, a decision of best split is made using impurity measures. The tree leaves are made up of the class labels which the data items have been group. Decision tree classification technique is performed in two phases: tree building and tree pruning.

Tree building is done in top-down manner. During this phase the tree is recursively partitioned till all the data items belong to the same class label. Tree pruning is done in a bottom-up fashion. It is used to improve the prediction and classification accuracy of the algorithm by minimizing over-fitting (noise or much detail in the training data set). Tree pruning has less tasking compared to the tree growth phase as the training data set is scanned only once. One of the strengths of Decision trees compared to other methods of induction is the ease with which they can be used for numeric as well as nonnumeric domains. The major advantage of Decision tree is that it can be easily mapped to rules.

2.3.3 IDE3

IDE3 (Iterative Dichotomiser 3) Decision tree algorithm was introduced in 1986 by Quinlan Ross[64]. It is based on Hunt’s algorithm and is serially implemented. Like other Decision tree algorithms the tree is constructed in two phases; tree growth and tree pruning. Data is sorted at every node during the tree building phase in-order to select the best splitting single attribute. IDE3 uses information gain measure in choosing the splitting attribute. It only accepts categorical attributes in building a tree model. IDE3 does not give accurate result when there is too-much noise or details in the training data set, thus a an intensive pre-processing of data is carried out before building a Decision tree model with IDE3.

2.3.4 C4.5

C4.5 algorithm is an improvement of IDE3 algorithm, developed by Quinlan Ross (1993). It is based on Hunt’s algorithm and also like IDE3, it is serially implemented. Pruning takes place in C4.5 by replacing the internal node with a leaf node thereby reducing the error rate[65]. Unlike IDE3, C4.5 accepts both continuous and categorical attributes in building the Decision tree. It has an enhanced method of tree pruning that reduces misclassification errors due to noise or too-much details in the training dataset.

Like IDE3 the data is sorted at every node of the tree in order to determine the best splitting attribute. It employs gain ratio impurity method to evaluate the splitting attribute[66].

* + 1. CART

CART (Classification and regression trees) was introduced by Breiman, (1984). It builds both classifications and regressions trees. The classification tree construction by CART is based on binary splitting of the attributes. It is also based on Hunt’s model of Decision tree construction and can be implemented serially. It uses gini index splitting measure in selecting the splitting attribute. Pruning is done in CART by using a portion of the training data set. CART uses both numeric and categorical attributes for building the Decision tree and has in-built features that deal with missing attributes[67]. CART is unique from other Hunt’s based algorithm as it is also use for regression analysis with the help of the regression trees. The regression analysis feature is used in forecasting a dependent variable (result) given a set of predictor variables over a given period of time. It uses many single variable splitting criteria like gini index, symgini etc and one multi-variable (linear combinations) in determining the best split point and data is sorted at every node to determine the best splitting point. The linear combination splitting criteria is used during regression analysis. SALFORD SYSTEMS implemented a version of CART called CART® using the original code of Breiman, (1984). CART® has enhanced features and capabilities that address the short-comings of CART giving rise to a modern Decision tree classifier with high classification and prediction accuracy.

* + 1. 2.3.6ROUGH SETS THEORY

Rough set theory[68-70] is a fairly new intelligent technique that has been applied to the medical domain, and is used for the discovery of data dependencies, evaluates the importance of attributes, discovers the patterns of data, reduces all redundant objects and attributes, and seeks the minimum subset of attributes. Some researchers had successfully adopted the Rough sets theory, for classifying stages of various medical complications[71]. Rough sets theory in medicine is an emerging field of high importance for providing prognosis and a deeper understanding of the classification of various diseases**.**

Bazan et.al.[72] compared Rough set-based methods, in particular dynamic reducts, with statistical methods, Neural networks, Decision trees and Decision rules, medical data on lymphography, breast cancer and other primary tumors. It was found that the error rates for Rough sets were fully comparable as well as often significantly lower than the ones obtained with other computational techniques. Carlin et al. presented an application of Rough sets for diagnosing suspected acute appendicitis complications.

Zhong et al.[73] applied Rough Sets with Heuristics (RSH) and Rough Sets with Boolean Reasoning (RSBR) for attribute selection and discretization of real-valued attributes. Komorowsk et al.[74] studied an application of rough sets to modeling prognostic power of cardiac tests. Hu et al.[75] developed two new algorithms to calculate core attributes and reducts for feature selection. These algorithms can be extensively applied to a wide range of real-life applications with very large data sets. Jensen et al.[76-78] developed the Quick reduct algorithm to compute a minimal reduct without exhaustively generating all possible subsets and also they developed Fuzzy-Rough attribute reduction with application to web categorization.

* + 1. 2.3.7 NEURAL NETWORKS IN MEDICINE

Artificial Neural Networks (ANN) is currently a 'hot' research area in medicine and it is believed that they will receive extensive application to biomedical systems in the next few years. At the moment, the research is mostly carried out on modeling parts of the human body and recognizing diseases from various scans (e.g. cardiograms, CAT scans, ultrasonic scans, etc.)[79]. Neural networks are ideal in recognizing diseases using scans since there is no need to provide a specific algorithm on how to identify the disease. Neural networks learn by example so the details of how to recognize the disease are not needed. What is needed is a set of examples that are representative of all the variations of the disease. The quantity of examples is not as important as the 'quantity'. The examples need to be selected very carefully if the system is to perform reliably and efficiently.

CHAPTER 3

DESIGN AND MODELLING OF MACHINE LEARNING METHODS

* 1. 3.1 OVERVIEW

The process of knowing, thinking, understanding, problem solving, judgment, processing of information regarding images, concepts, words, rules, symbols and creativity is termed as cognition. Cognition is that operation of mind by means of which, we become aware of our surroundings, objects and thoughts. Cognitive disorders are mental disorders characterized by impaired cognitive abilities and daily functioning in which biological causation is either known or presumed. Cognitive disorders involve disturbance in thinking or memory that represent a marked change from the individual’s prior level of functioning[132].Cognitive disorders are not psychologically based; they are caused by physical or mental conditions or drug use or withdrawal that affects the functioning brain. In some cases the special cause of the cognitive disorders are identified. Cognitive disorders arise when the brain is either damaged or impaired in its ability to function due to injury, illness, exposure to toxins or use or abuse of psychoactive drugs..

* 1. 3.2 ARCHITECTURE AND MODELING

The traditional method of assessing the mental status of the dementia patients is done using two important screening tools i.e., MMSE and FAQ, which are recommended by the Agency for health care policy research. In the current study Machine learning method has been applied along with the two screening tools in order to assess the cognitive state of the dementia patients. The architecture and modeling of the present work has been presented in Figure 4.1.The steps involved were collection of records, preprocessing, feature selection and finally applying of various classifier techniques.

This model begins with the collection of patients records. The patients records were collected from various places. The initial visits of 860 subjects seen either as control or as patients from the National Institute on aging which consists of patients and caregiver interviews. The National Institute of Aging, a component of the National Institutes of Health, USA supports 28 Alzheimer's disease centers across the country. The set of data was used for building the model. In order to test the model around 90 subjects were collected from various neurocenters across north Karnataka. In the study most of the collected data was free from outliers, noise and missing data. Some inconsistencies were found in the patients reports and these consistencies were corrected manually after consulting with some well-known neuropsychiatrists.

Features Selection Cognitive Tests: MMSE and FAQ

Classifier C4.5 Rules

Naïve Bayes Neural Networks Random Forest

Collection of Patients Records

Figure 3.1: Architecture for Classifying Cognitive State of Dementia Patients

Processing of Records

* + 1. 3.3 FEATURE SELECTION

In the current system the functional and cognitive abilities of the patients were extracted by conducting some memory examinations such as MMSE and FAQ, these tests could be used to systematically and thoroughly assess mental status.

The MMSE is the standard for assessing cognitive function; it is a validated test that is highly sensitive to moderate and severe dementia. The MMSE tests and scores six domains: (1) orientation, (2) registration (3) attention and calculation (4) recall

(5) language (6) constructional ability (Table 4.1).

3.4 MACHINE LEARNING METHODS

Machine Learning and Knowledge Discovery from Databases (KDD)[139] are increasingly being applied in health care to build models, develop practice guidelines or refine guidelines for better medical decision making. They differ from traditional approaches by generating domain models such as Decision trees, Decision rules, graphs etc. from data.

ML methods attempt to learn a description that best separates the different classes of dementia state. As input, each of the initial visits is represented with the attributes described above. The output of each of the learning methods is a representation of the dementia state that can be applied to classify patients with an unknown dementia status. Each learning method applies a different search technique and concept representation to describe the possible outcomes of the diagnosis. In the proposed system each of the four ML methods for the classification of the dementia state have been applied.

* + 1. CLASSIFICATION AND REGRESSION TREES (CART)

In the present study Classification and Regression Trees (CART) classification algorithm have been adopted[140]. Classification tree analysis predicts the outcome of the class to which the data belongs. Regression tree analysis predicts the outcome which can be considered a real number. CART has been applied in number of applications in the medical domain[141-143]. One of the advantages of using classification trees is their ability to provide easy to understand classification rules. Each node of a classification tree is a rule. The only exception to this would be in cases where the tree is very large and in such cases there may need to be a more specific focus on pruning required to optimize the tree size. Trees are easy off the shelf classifiers that require no variable transformation. CART builds the tree by recursively splitting the variable space based on the impurity of the variables to determine the split till the termination condition is met. The gini impurity determines how often a randomly chosen element from the set would be incorrectly labeled if it were randomly labeled according to the distribution of labels in the subset. The following is a pseudo procedure[144]. The CART algorithm implemented in Weka uses minimal cost-complexity pruning. The Figure 2 shows the classification tree generated by Weka

* + 1. DECISION TREE

Decision tree induction represents a simple and powerful method of classification which generates a tree and a set of rules, representing the model of different classes, from a given dataset[145]. Decision Tree (DT) is a flow chart like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test and each leaf node represents the class. The top most node in a tree is the root node. For DT induction, ID3 algorithm and its successor C4.5 algorithm by J. R. Quinlan[146] are widely used. Algorithm (java implementation of C4.5) is used in this paper for Decision tree induction. One of the strengths of Decision trees compared to other methods is the ease with which they can be used for numeric as well as nonnumeric domains. Another advantage of decision tree is that it can be easily mapped to rules. The classical DT induction algorithm i.e. C4.5 algorithm by Quinlan (1993) is presented below:

Let the training data is a set *S* = *s*1, *s*2 ... of already classified samples. Each sample *si* = *x1, x2* ... is a vector where *x1, x2* represent attributes or features of the sample. The training data is augmented with a vector *C* = *c*1, *c*2... where *c*1, *c*2...represent the class that each sample belongs to. C4.5 uses the fact that each attribute of the data can be used to make a decision that splits the data into smaller subsets.

C4.5 examines the normalized Information Gain (difference in entropy) that results from choosing an attribute for splitting the data. Entropy(*S*) can be thought of as a measure of how random the class distribution is in *S*. Information gain is a measure given to an attribute a. Attribute a can separate *S* into subsets *Sa*1, *Sa*2, *Sa*3... *San*. The information gain of a is Entropy(*S*) – Entropy(*Sa*1) – Entropy(*Sa*2) – ... – Entropy(*San*). Information gain is then normalized by multiplying the entropy of each attribute choice by the proportion of attribute values that have that choice. The attribute with the highest normalized information gain is the one used to make the decision. The algorithm recurs on the smaller sub lists.

* + 1. NAÏVE BAYES

Bayesian classifiers are statistical classifiers. They can predict class membership probabilities, such as the probability that a given sample belongs to a particular class. Bayesian classifier is based on Bayes’ theorem. Naive bayesian classifiers assume that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is called class conditional independence. It is made to simplify the computation involved and, in this sense, is considered “naïve”.

Let *X* = {*x1, x2. . . xn*} be a sample, whose components represent values made on a set of n attributes. In Bayesian terms, *X* is considered “evidence”. Let *H* be some hypothesis, such as that the data *X* belongs to a specific class *C*. For classification problems, our goal is to determine P(*H*|*X*), the probability that the hypothesis *H* holds given the "evidence", (i.e. the observed data sample *X*). In other words, we are looking for the probability that sample *X* belongs to class *C*, given that we know the attribute description of *X*. P(*H*|*X*) is the a posteriori probability of *H* conditioned on P(*H*|*X*) is the a posteriori probability of *H* conditioned on *X*.

Fox example, suppose our data samples have attributes: orientation and registration age and income, and that sample *X* is a 72-year-old patient with an MMSE score 17. Suppose that *H* is the hypothesis that the patient is having cognition problem. Then P(*H*|*X*) is the probability that the patient *X* will be suffering from dementia. The a posteriori probability P(*H*|*X*) is based on more information (about the customer) than the a priori probability, P(*H*), which is independent of *X*. Similarly, P(*X*|*H*) is the a posteriori probability of *X* conditioned on *H*. P(*X*) is the a priori probability of *X*. According to Bayes’ theorem, the probability that we want to compute P(*H|X*) can be expressed in terms of probabilities P(*H*), P(*X*|*H*), and P(*X)* as and these probabilities may be estimated from the given data.

P(*H*|*X*) =

P(*X*|*H*) P(*H*) P(*X*)

The Naive bayesian classifier works as follows:

1.Let *T* be a training set of samples, each with their class labels. There are k classes, *C1,C2, . . . ,Ck*. Each sample is represented by an n-dimensional vector, *X* = {*x1, x2. . . xn*}, depicting n measured values of the *n* attributes, *A1,A2, . . . , An*, respectively.

2.Given a sample *X*, the classifier will predict that *X* belongs to the class having the highest a posteriori probability, conditioned on *X*. That is *X* is predicted to belong to the class *Ci* if and only if

P(*Ci*|*X*) > P(*Cj* |*X*) for 1< *j*<*m*,

* + 1. NEURAL NETWORK

Multilayer perceptron are feedforward Neural networks trained with the standard Backpropagation algorithm. They are supervised networks so they require a desired response to be trained. They learn how to transform input data into a desired response, so they are widely used for pattern classification. With one or two hidden layers, they can approximate virtually any input-output map. They have been shown to approximate the performance of optimal statistical classifiers in difficult problems. Most Neural networks applications involve MLPs. This is perhaps the most popular network architecture in use today,

In a Multilayer perceptron topology neurons are grouped into distinct layers as depicted in Figure 3. Output of each layer is connected to input of nodes in the following layer. Inputs of the first layer (input layer) are the inputs to the network while the outputs of the last layer, form the output of the network. A Multilayer perceptron is especially useful for approximating a classification function that maps input vector (*x1, x2, ... xn*) to one or more classes *C1, C2,...Cm*.*.* By optimizing weights and thresholds for all nodes, the network can represent a wide range of classification functions. Optimizing the weights can be done by supervised learning, where the network learns from the large number of examples. Examples are usually provided one at a time. For each example the actual vector is computed and compared to the desired output. Then, weights and thresholds are adjusted, proportional to their contribution to the error made at the respective output. One of the most used methods is the Backpropagation method, in which in the iterative manner, the errors are propagated (error = the difference between desired output and the output of actual ANN) into the lower layers, to be used for the adaptation of weights.

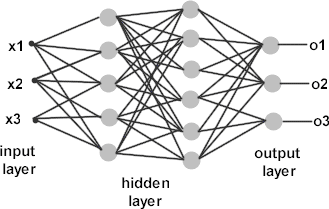


Figure 3.2: Architecture of Multilayer Perceptron

**CHAPTER 4**

**IMPLEMENTATION**

**4.1 Programming language selection**

Python is the programming language, it is widely used open source general purpose scripting language that is especially suited for web development and can be embedded into HTML.

MySQL is used with PHP as back-end tool, where MySQL is the popular online database and can be interfaced very well with PHP.

**4.2 User Interface and Functionality implementation**

The user interface is how the system interacts with the user. If the user enters some data, the system must fetches the input and it must make some respond through displaying a message on the screen relating to the input or opening a new page according to the input entered by the user.

To check the validity of data that are entered by the user, there are some functions used in this project. It fetches the data and checks whether the data is valid or not.

**CHAPTER 5**

**TESTING**

|  |  |
| --- | --- |
| **TEST 1** | |
| Name | Signup |
| Action | Signup as doctor or patient or admin to website |
| Output | It displays as User Created Successfully |
| Result | **Pass** |

|  |  |
| --- | --- |
| **TEST 2** | |
| Name | Signin |
| Action | Patient or Doctor or Admin logins to web through his credentials |
| Output | It gets logged in to user portal to book consultation or view feedback from patients |
| Result | **Pass** |

|  |  |
| --- | --- |
| **TEST 3** | |
| Name | Predict Disease |
| Action | Clicking on Predict button |
| Output | It displaysthe Disease with probability rate depending on symptoms given |
| Result | **Pass** |

**CHAPTER 6**

**SNAPSHOTS**

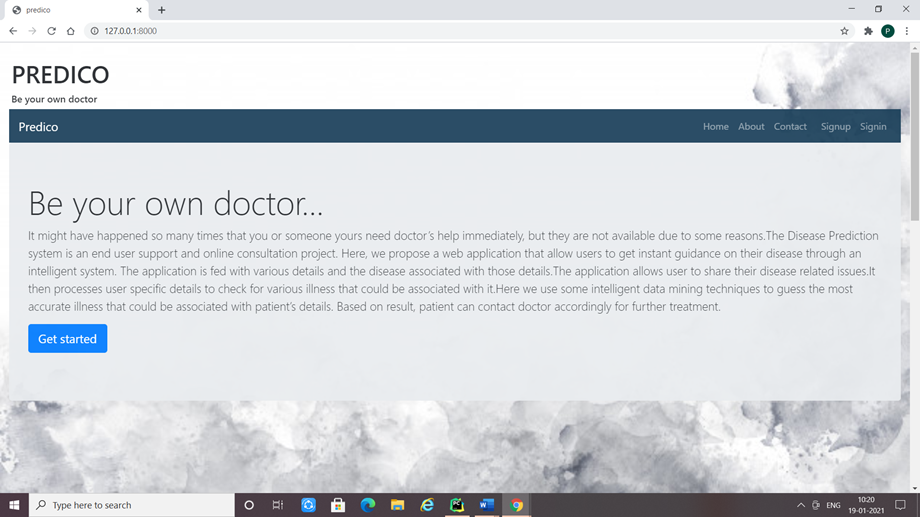
****

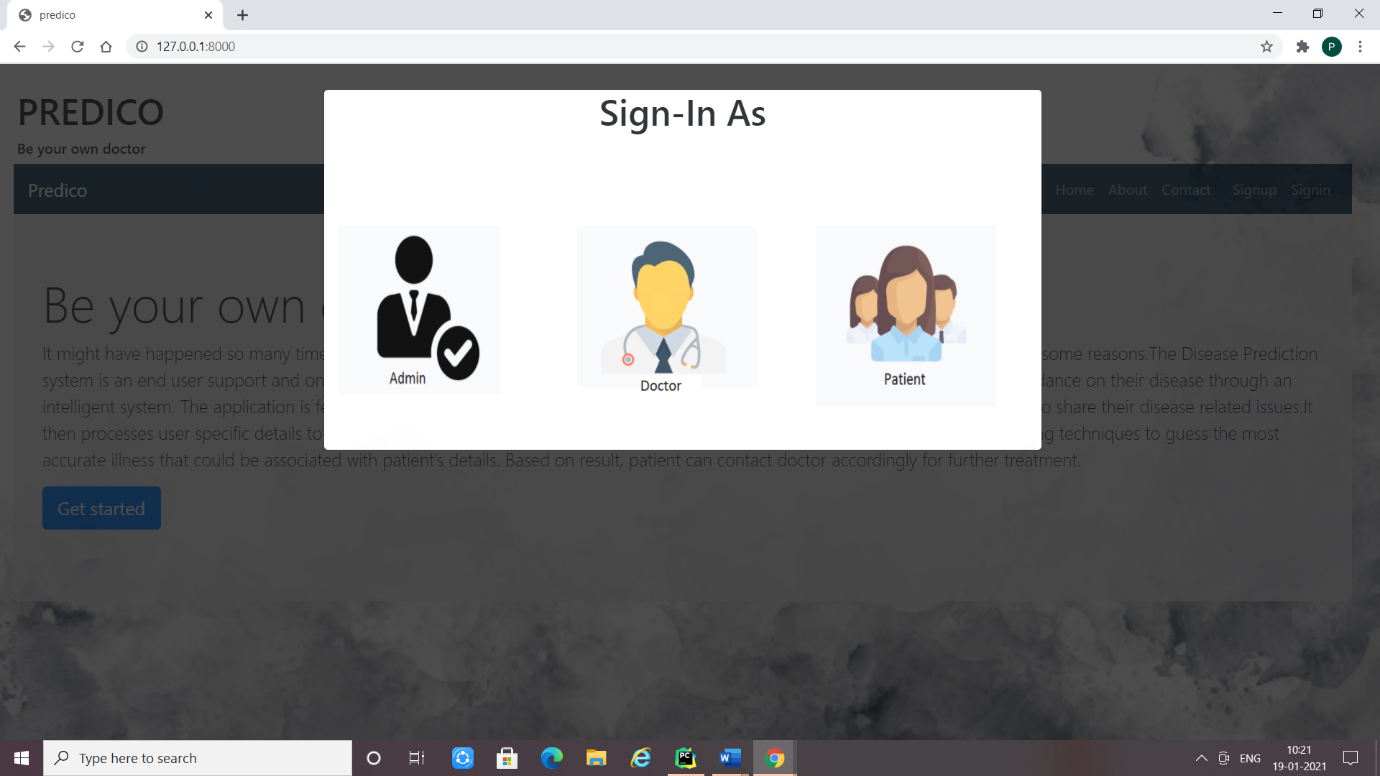
Fig 6.1 Home page

Fig 6.2 About us

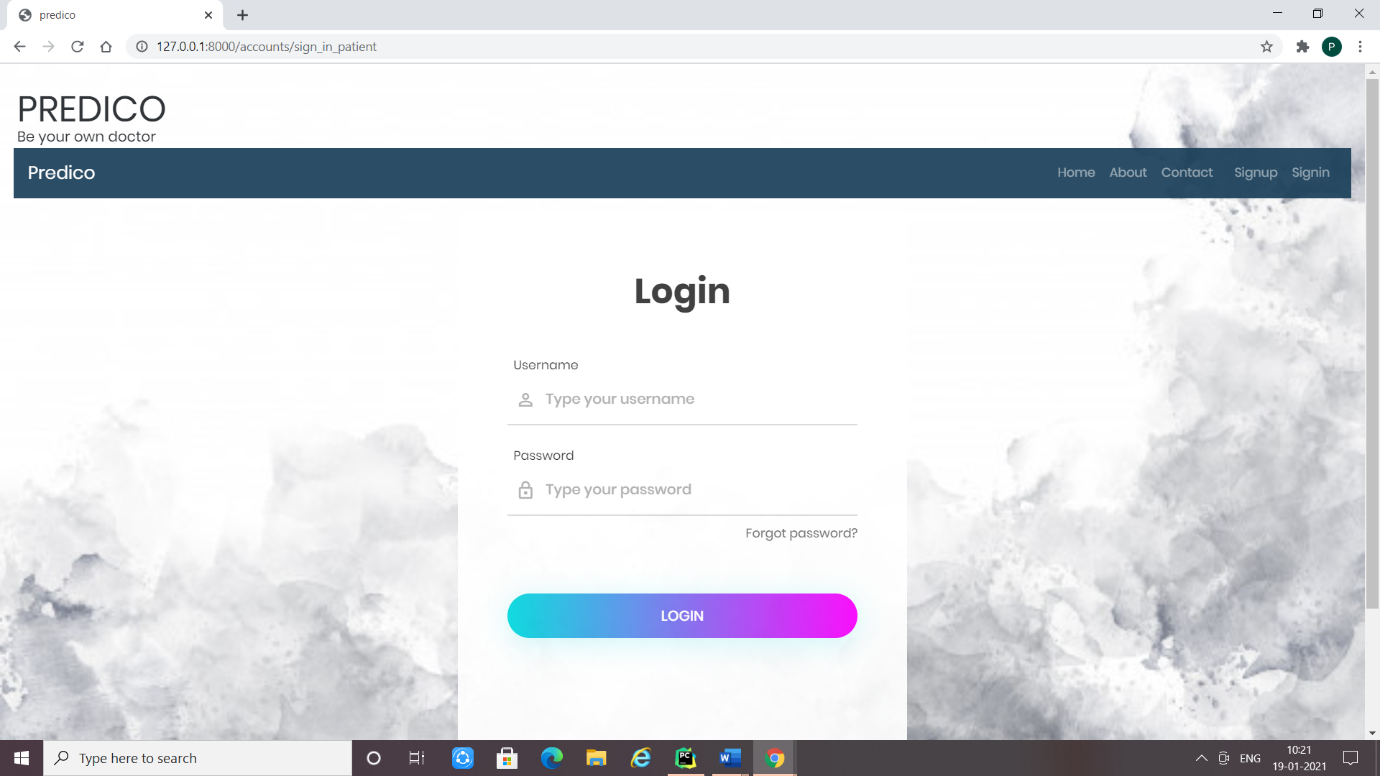


Fig 6.3 login page

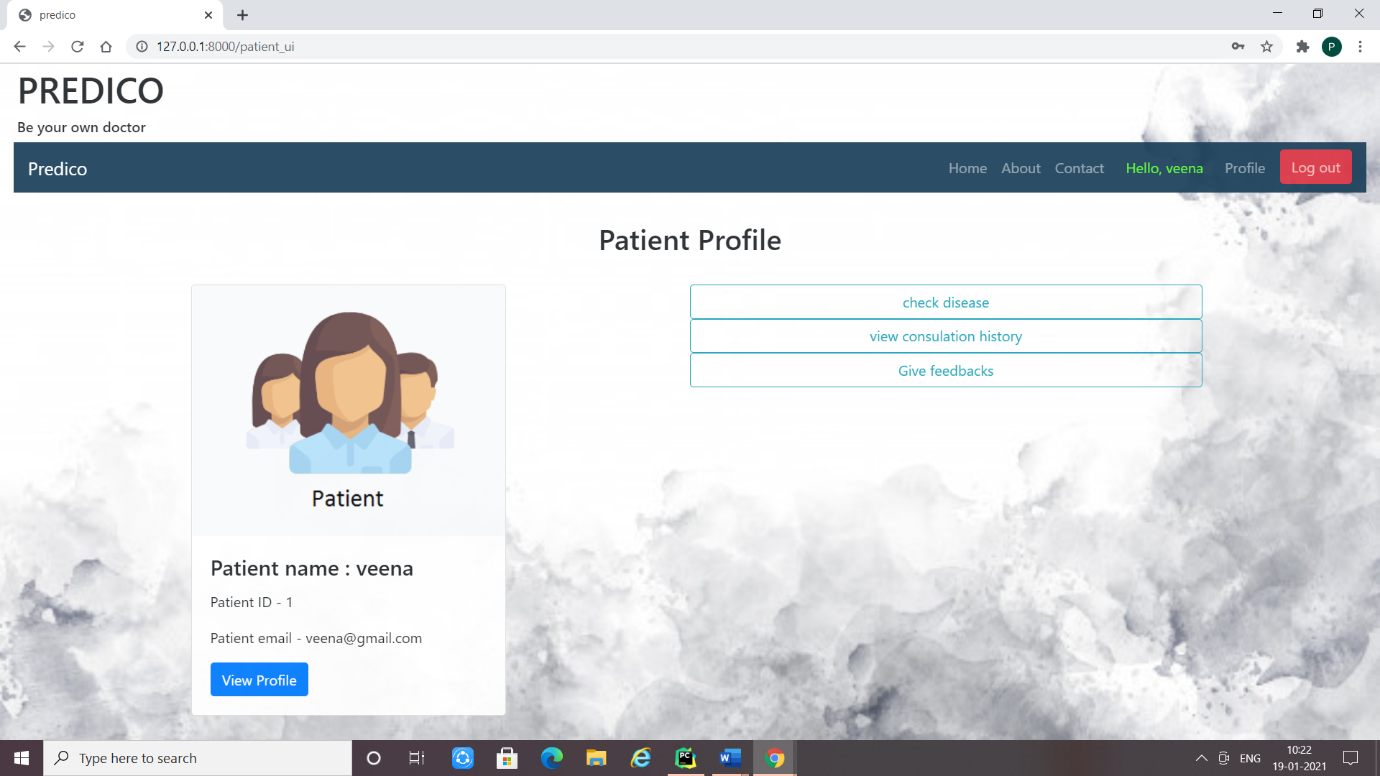


Fig 6.4 patient profile

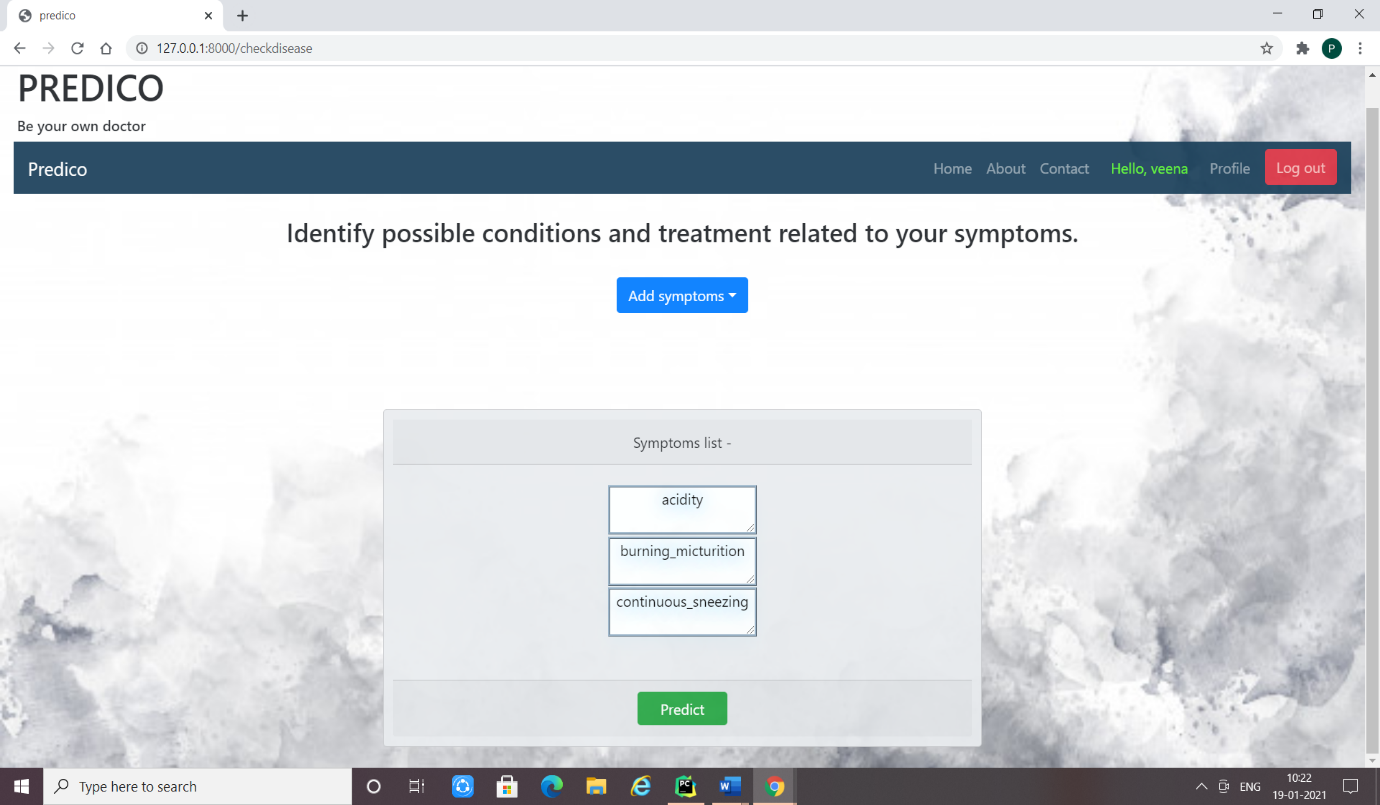


Fig 6.5 check diseases

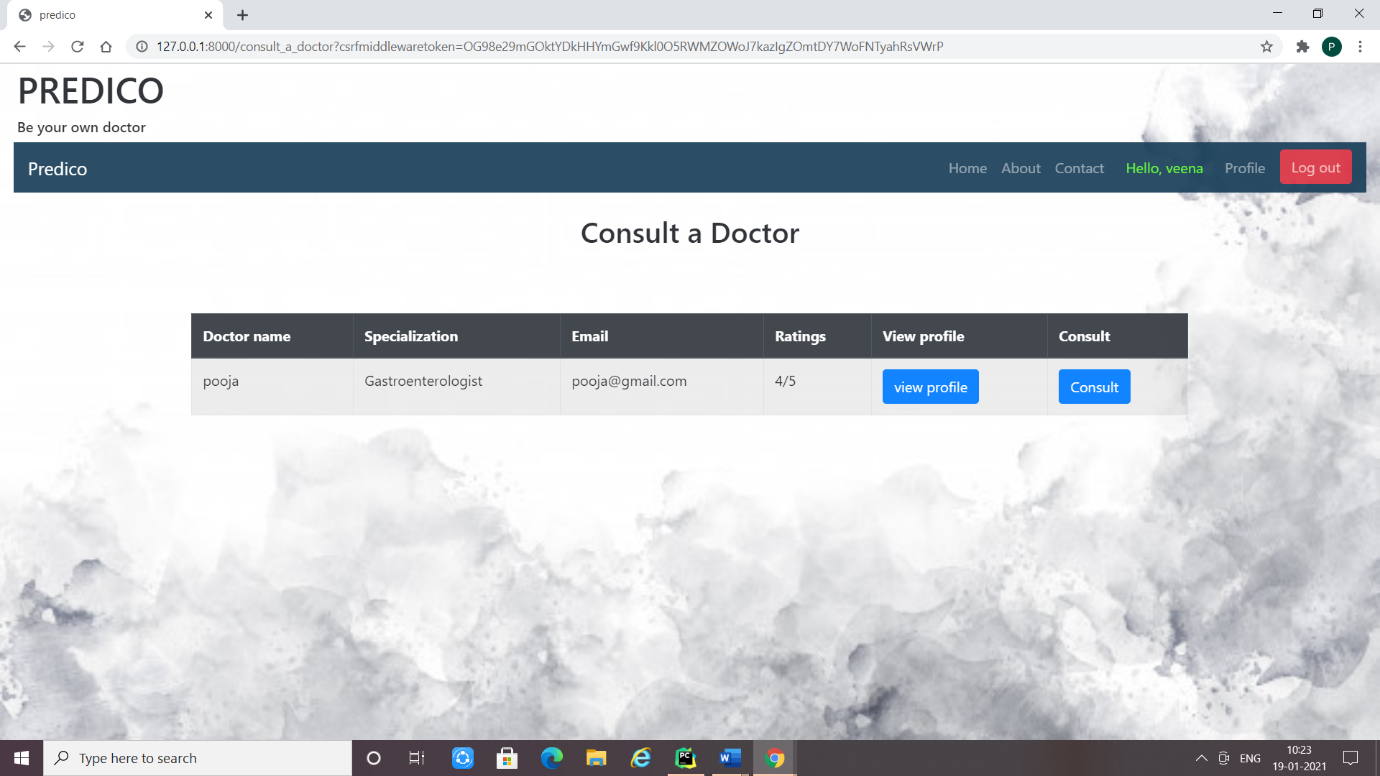


Fig 6.6 Doctor Consult

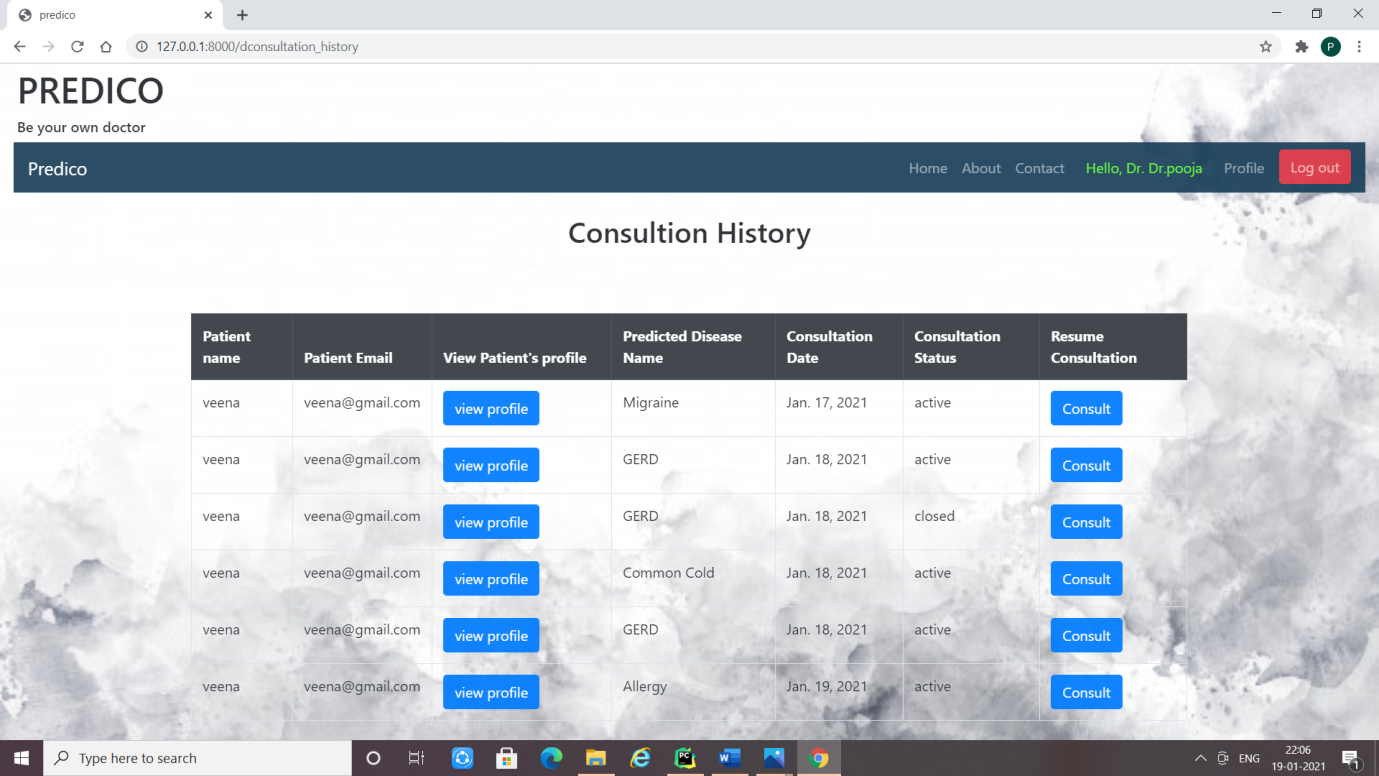


Fig 6.7 Consultion History

**CONCLUSION**

This project has made us aware of the immense capabilities and the various uses of HTML, PHP, CSS, MySQL and APACHE server both individually and combined.

The disease prediction system have 3 users such as doctor, patient and admin. Each user of the system are authenticated by the system.There is a role based access to the system.The systemallows the patient to give symptoms and according to thosesymptoms the system will predict a disease. The system suggests doctors for predicted diseases. It allows online consultation for patients.

**REFERENCES**

1.Min chen ,yixueHao,kaiHwang,Fellow,IEEE,Lu Wang, and Lin Wang “Disease Prediction by Machine learning Over Big Data from Healthcare Communities”[2017].

2.Mr.ChalaBeyene, Prof.Poojakamat, “Survey on Prediction and Analysis the occurance of Heart Disease Using Mining Techniques”,International Journal of pure and Applied mathematics,2018.