Prediction of Pediatric Irritable Bowel Syndrome using Machine Learning Ensemble Approach

A Thesis

submitted in partial fulfilment of the requirements for the award of the degree of

Master of Engineering

in

Computer Science and Engineering

by

Ashish Jat

(Roll No: 801532006)

Under the supervision of

Dr. Maninder Kaur

(Assistant Professor)

Computer science and Engineering Department

THAPAR UNIVERSITY

PATIALA-147004, PUNJAB, INDIA

June 2017

Certificate

I hereby certify that the work which is being presented in the thesis entitled,"*Prediction of Pediatric Irritable Bowel Syndrome using Machine Learning Ensemble Approach*", in partial fulfillment of the requirements for the award of degree of Master of Technology in Computer Science and Engineering submitted in Computer *Science and Engineering Department* of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of Dr.Maninder Kaur and refers other researcher's work which are duly listed in the reference section. The matter presented in the thesis has not been submitted for the award of any other degree of this or any other University.

(Ashish Jat)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

(Dr. Maninder Kaur)

Assistant Professor, CSE

Department

Acknowledgement

First of all, I would like to thank the Almighty, who has always guided me to work on the right path of the life. This work would not have been possible withoutthe encouragement and able guidance of my supervisor **Dr. Maninder Kaur**. I thank my supervisor for their time, patience, discussions and valuable comments. Their enthusiasm and optimism made this experience both rewarding and enjoyable.

I am equally grateful to **Dr. Maninder Singh**, Associate Professor, and Head, Computer Science & Engineering Department, a nice person, an excellent teacher and a well–credited researcher, who always encouraged me to keep going with work and always advised me with his invaluable suggestions. I will be failing in my duty if I don't express my gratitude to **Dr. S.S. Bhatia**, Senior Professor and Dean of Academic Affairs, Thapar University, for making provisions of infrastructure such as library facilities, computer labs equipped with net facilities, immensely useful for the learners to equip themselves with the latest in the field.I am also thankful to the entire faculty and staff members of Computer Science and Engineering Department for their direct-indirect help, cooperation, love, and affection ,which made my stay at Thapar University memorable.

Last but not least, I would like to thank my family whom I dearly miss and without whose blessings none of this would have been possible. To my parents, I own thanks for their wonderful love and encouragement. I would also like to thank my brother since he insisted that I should do so. I would also like to thank my close friends for their constant support.

Date: June 2017

Place:Thapar University, Patiala

(Ashish Jat)

Abstract

Machine learning Ensembling approach has the potential to resolve Irritable Bowel Syndrome(IBS) problem. Machine learning techniques have numerous benefits that include high flexibility and power, lack of parametric assumptions, etc. The researchers do not properly understand the causes of the IBS. The researchers found that the IBS caused due to the combination of the physical and the mental health problems.Ensemble methods combine the predictions from the various machine learning algorithms which use these predictions as inputs for the second-level learning models.

This research focuses on detection of Irritable Bowel Syndrome (IBS) using machine learning ensemble approach. The experimental analysis is performed using various machine learning models: Support vector machines (SVM), Neural Network, Linear Regression, Random Forest, Decision tree, AdaBoost (Adaptive Boosting), Naive Bayes, Boosted tree, Multilayerperceptrone, and Binary Discriminate analysis. The data was collected from the Website of UMASS Medical School. The collected data wasof the pediatric patients. The data was used to predict the presence of IBS in pediatric patients. In our research, we ensemble ten different models to build a new model having high accuracy to predict a pediatric patient is IBS or not. The implementation of proposed ensemble model was done in R language. The RRF model was used for feature selection task. We used R language for the implementation of the proposed ensemble model. The RRF model was used for feature selection task. Preliminary results of the experiment show that our model is 93.32% accurate in predicting whether a pediatric patient is IBS positive or not.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Irritable Bowel Syndrome (IBS) is a disease of human small intestine and colon. The symptoms of IBS are non-specific. Therefore the diagnosis could be delayed because of the required confirmation from invasive colonoscopy. This delay leads to the poor treatment results[1]. Many existing systems used machine learning models to determine the presence of IBS. Irritable Bowel Syndrome (IBS) is a common type of disorder which affects the large intestine of the body. It causes abdominal pain, cramping, diarrhea, bloating gas, and constipation. It is a chronic condition which manages in the long term. Some people have severe symptoms and signs of irritable bowel syndrome. People can control these symptoms by managing their stress, diet, and lifestyle while others will need counseling and medication [6].

1.1. Irritable bowel syndrome

Irritable bowel syndrome abbreviated as IBS is one of the common disorder that impacts the larger intestine of a human being. Cramping, bloating, diarrhea, gas, abdominal pain and constipation are some major health issues that occur as a consequence of Irritable bowel syndrome (IBS). It is common functional disorders. The word function disorder here defines the problem with the proper functioning of any part of the body. It does not cause any abnormality in the structure, but likely to disrupt the function of a part of the body. In simple words, even if we see the gut under the microscope, we will not see any structural problem in the gut. All parts would look working normally [1]. But it would cease to function appropriately.

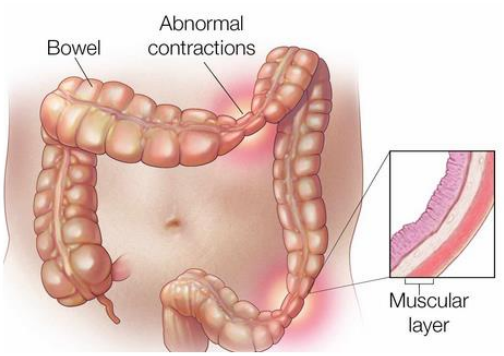


Figure 1: Irritable Bowel Syndrome (IBS) [13]

IBS is a functional intestinal disorder which is caused by the abnormal functioning of the gastrointestinal tract (GI- tract). It is considered as the common disorder which occurs in the large intestine. The women are mostly affected by this syndrome as compared to the men.IBS affects the adults also. IBS is not a life-threatening illness; it occurs due to the unsubstantial health care and burden on the mind. IBS disease occurs due to some physical irregularities such as the infection and the metabolic disturbances which remain in the patients for many years. Sometimes the examiners also not find the exact problem regarding this disease and it remained unclear. This syndrome is now considered as the biological and the neurological bowel syndrome. IBS often deliberated as the spastic disease which mainly affects the nervous system of the person. IBS is the short- tempered colon which affects the intestines inside the human body [12].

IBS is a long-term disorder which mainly affects the intestines inside the human body due to the disturbance occurs in the human body such as the intestinal problems and the sensational problems. It creates the unconscious activities which are regulated by the brain. So, IBS is also considered as the Brain-gut disorder. The disturbances which occur inside the human body can harvest the symptoms of the abdominal pain, bloating and the gaseousness senses and the person also suffer from constipation[6]. The other factors which affect the human body are visceral hypersensitivity and the other psychological factors.

There is only limited information available about the IBS in the children. The report based on research states that normally 10 to 20 percent children suffered from the IBS. The children who suffer from the problem faced lots of symptoms like abdominal pain, indigestion, and abdominal migraine. The study also shows the one thing that the IBS affects both boys and the girls equally.

1.1.1. Causes of the IBS

The causes of the IBS which affects the children are discussed below:

1.1.1.1. Brain- Gut signal problems

The signals between the brain and the nerves of the small and the large intestines also called the instinctive. If there is some problem occurs in the brain, the

signals relating to the small and the large intestine can cause the IBS symptoms. The symptoms which are shown are changes in the bowel habits and creates the bloating in the stomach [6].

1.1.1.2. GI motor problems

The normal motility in the children who has the IBS disease may be present in the colons of the large intestines. The slow motility can lead them to the problems which are relating to the stomach like constipation and the fast motility lead children to the serious diseases like Diarrhoea [43]. Sometimes it can also create the muscle shrinkages which also cause abdominal pain. Some children also face the problem of hyperactivity with the increase contractions between the bowel movements. These bowel movements can also give a response to the stress or the eating problems [1].

1.1.1.3. Hypersensitivity

People who suffer from the IBS have the greater hypersensitivity in the abdominal pain as compared to those who don't have the IBS. People who are suffering from the IBS have the different rectal tone and the rectal motor response after eating their meals. They suffer from the problem like abdominal pain and the bloating in the stomach [12].

1.1.1.4. Mental health problems

IBS disease sometimes becomes adangerous disease, and it can create some mental and the psychological problems such as nervousness and depression. These psychological problems can lead patients to some serious health problems.

1.1.1.5. Bacterial gastroenteritis

The children who are suffering from the bacterial gastroenteritis which occurred due to the infection in the stomach can also cause the IBS in the children's. The scientists have shown the connection between the gastroenteritis and the IBS in the adults only not in the children. The gastroenteritis can lead to the disease IBS in the children and shows the worst result inside the body [41].

1.1.1.6. Small intestinal bacterial overgrowth *(*SIBO*)*

The bacteria inside the stomach normally live in the small intestine. Due to increasing number of bacteria in the small intestine can cause the problems of the IBS

in the children. These bacteria sometimes produce gas in the stomach and may cause diarrhea and weight loss. Some researchers express that there is a relation between the SIBO and the IBS. The SIBO may lead to the IBS disease. But some studies have also found the antibiotics for the treating of the IBS. So, there is a need to find out the connection of the SIBO with the IBS [12].

1.1.1.7. Genetics

Sometimes the IBS might be caused by the genetic problems as it runs in the family from one generation to the other generation. The studies also have shown that the IBS is more common in the people whose family members have the same problem relating to the IBS [44].

1.1.2. Symptoms of IBS

IBS consists of a large list of symptoms. The symptoms of the IBS include the abdominal problem or the discomfort in the bowel habits. The person with IBS can have discomfort and pain that might not disable for a long time [12].The primary symptoms of IBS include the following:

1.1.2.1. Bloating and Swelling

An individual with IBS might realize bloating and swelling of the stomach time to time [42].

1.1.2.2. Abdominal Pain

The person with IBS will realize pain and discomfort in avarious part of the stomach. Generally, the stomach pain comes and goes [6].

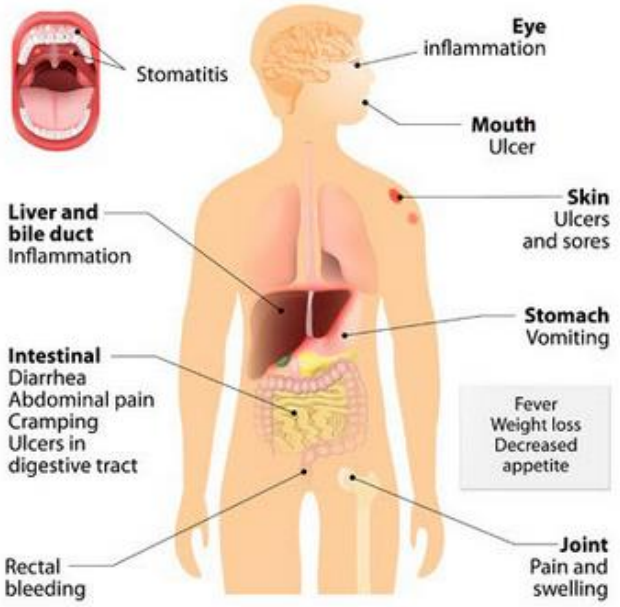


Figure 2: Symptoms of IBS [14]

1.1.2.3. Diarrhoea

Diarrhea is the main symptom which occurs in the IBS. It commonly occurs in the children and the adults. They face the problem having the loose and the watery stool to three or four times in the days as compared to the other days. The person is always feeling the urgency to have a bowel movement [12].

1.1.2.4. Constipation

The other main symptom of the IBS is constipation. The person who is suffering from this problem may have the hard and the dry stools. They suffer from the stress and improperbowel movement [45].

1.1.3. How IBS can be treated

The researchers have found that there is no treatment which helps to cure the IBS. But the symptoms can be treated by taking the following points in the mind that are:

- By changing the habits of eating all day inconstantly. Patients must take the proper diet and the nutrition in their daily routine. So that they cannot face the problem that occurs due to the IBS [2].

- The IBS can be cured by doing the exercise and the meditation daily. If the adults and the children do exercise and the meditation daily, then they get relief from the problems that occurred because of the IBS like abdominal pain and the bloating.
- If the probiotics were given to the children, then they may get relief from the IBS disease easily.
- Therapies mustbe given to the patient who is suffering from the mental health problems which can cause stress and depression in the patient.

1.1.4. Properties of IBS considered in this research

Irritable bowel syndrome **(**IBS**)** is associated with the morbidity in adolescents and children whose available treatment is limited**.** The majority of the people suffering from Irritable bowel syndrome will be deficient from the Vitamin D which affect their quality of life**.** Vitamin D deficiency is extremely widespread in the IBS patients which lead in the therapeutic implications**.**

Erythrocyte Sedimentation Rate is a test used to measure the amount the inflammation in the human body**.** It monitors and detects the inflammatory diseases or cancer**.** ESR is used in the diagnosis of Irritable bowel syndrome to remove the causes of the inflammatory of abnormal digestive symptoms**.**

Albumin Alanine Transaminase is a test which measures the amount of enzyme in the blood and found in the liver but also found in small amount in kidney, muscles, heart, pancreas**.** It is measured that liver is damaged or diseased**.** If the level of ALT is increased, then it may cause the liver damage which increases the chances of IBS in the body**.**

BMI is the body mass index calculated by dividing the weight of a person to its height squared**.** It is used to measure the health and risk of chronic diseases**.** It determines the person is underweight, obese, overweight, or a heathy weight respect to height**.** If the weight increases or decreases out of the healthy weigh range, then it can increase the risk of Irritable bowel syndrome**.**

1.2. Introduction to the machine learning

Machine learning is a technique which uses computers to use the example which is related to the data and the past experiences that help to solve the problems. Nowadays, in businesses, many applications come into the existence that helps to analyze the sales which have donein the past [5]. The machine learning also helps to predict the behavior of the customers and then optimize the robot behaviors in them so that the given task can be completed by using the minimum resources and provides the information in the form of informatics data [3].The other concept in the machine learning is related to find the patterns in the data. With the help of these patterns, we can predict the futureaspects and takethe decisions according to these patterns [4].

For example, the learning and the reading behavior of a person can be measured with the help of machine learning. With the help of machine learning techniques, one can identify that on which chapter the reader should devote the more time. It also helps to analyze that on which chapter the reader has more interest.Machine learning concepts also help to identify the activities and the time spent by the reader on the learning content. By analyzing the learning patterns of the readers, the author could know on which area the people will take the more interest and what must be improved and on what thing he must focus on in the future.

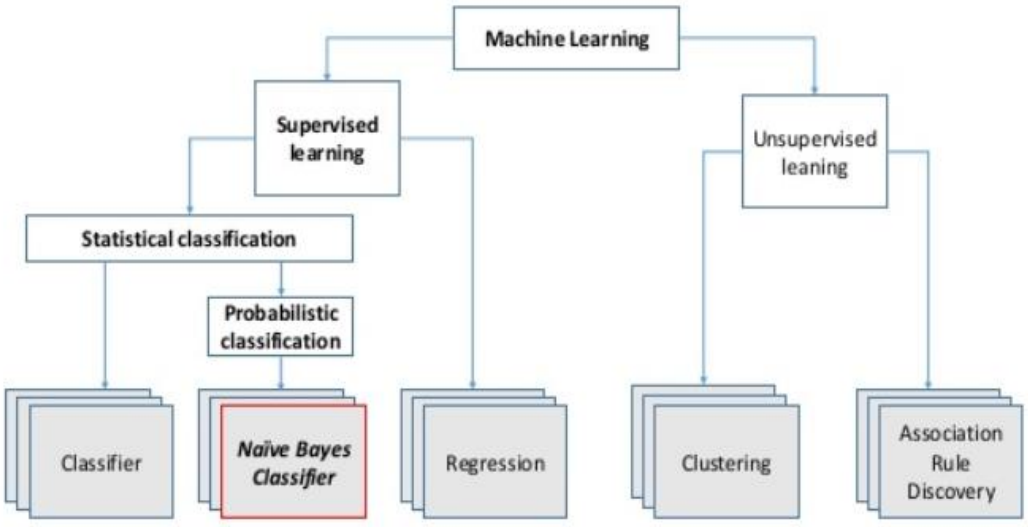


Figure 3: Some Machine Learning algorithm types

Another example of theuse of machine learning relates to finding frauds in the Internet Banking. If there is a continuous problem regarding transferring the funds from the internet banking. But we could not find out the pattern from which we can find the loopholes and the necessary information from which the fraud attacks the network. Machine learning techniques could help to discover the patterns, and it also helps to predict the results from that patterns. So, machine learning plays a vital role in the data mining, image processing and in the language processing. The machine learning cannot provide the correct data always,but it provides the analytical results which are based on the historical data to take the decisions in the future.

1.2.1. Techniques of machine learning

There are several techniques of Machine Learning. Some of them are discussed in this section.

1.2.1.1. Support Vector Machine*(*SVM*)*

Support VectorMachine(SVM) helps to classify data by creating dimensional hyperplane. Support Vector Machines are related to the neural networks. The SVM model uses the sigmoid kernel function to separates the data into the two categories. It gives the data in the two layers of perception which occurs in the neural networks [3]. This model also enables the multilayer perceptron which also occurred in the neural networks. By using the kernel function, the SVM provides the training method for the polynomial, radiated base function and the multi-layer perceptron [10]. By taking all these functions the weight of the network is taken to solve the quadric programming problem with the linear constructions. It does not solve the non-convex and the unrestrained minimization problems in the standard neural network.

In the SVM model, the prognosticator variable is known as the attribute. The transformed attribute in the SVM model is used to define the hyperplanewhich is known as the feature. The set of features which describes only one case is called a vector in the SVM model.

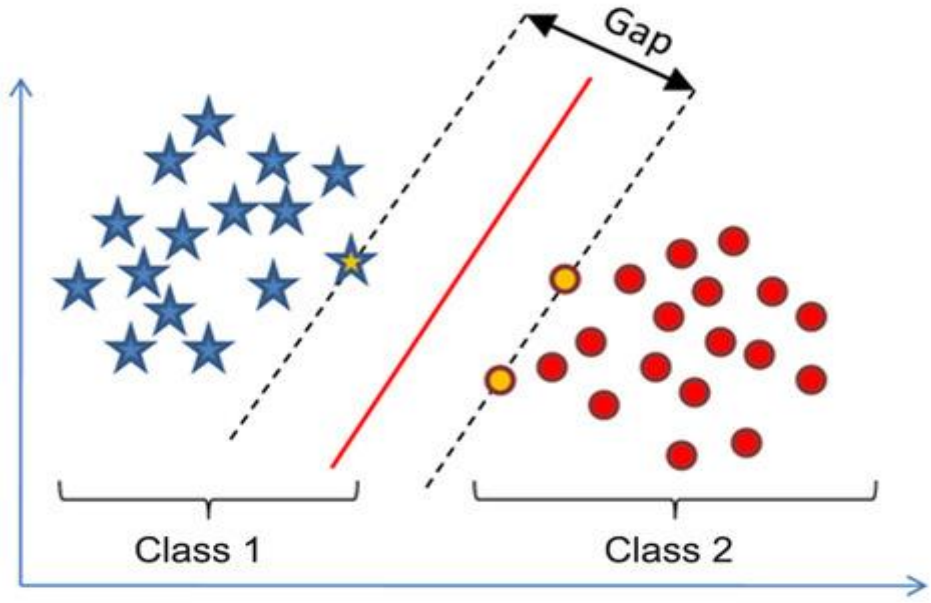


Figure 4: Basic Concept of SVM [16]

The main goal of the SVM model is to find the best hyperplane which isolates the clusters into two categories. It divides the cluster in that way that the one category of the objective variable is on one side of the hyperplane and the cases with the other category will place on the other size of the plane. In the SVM model,the vectors which place near the hyperplaneare known as the support vectors.

1.2.1.2. Neural Network

Neural network performs the number of regression which classifies the task into eachcommon network platforms thatperform only one function in the neural network. But in majority cases, these networks have single output variables which classify the problems into the number of output units. If one single network relates to the multiple output variables, then the variables inside the neural network suffer from the cross talk. So, the best solution is to put the variables into the separate network in each output unit. After putting the variables into the separate network then combine them ensemble so that they run as a unit in the neural network. Neural networks are as the multilayer perceptron [8].

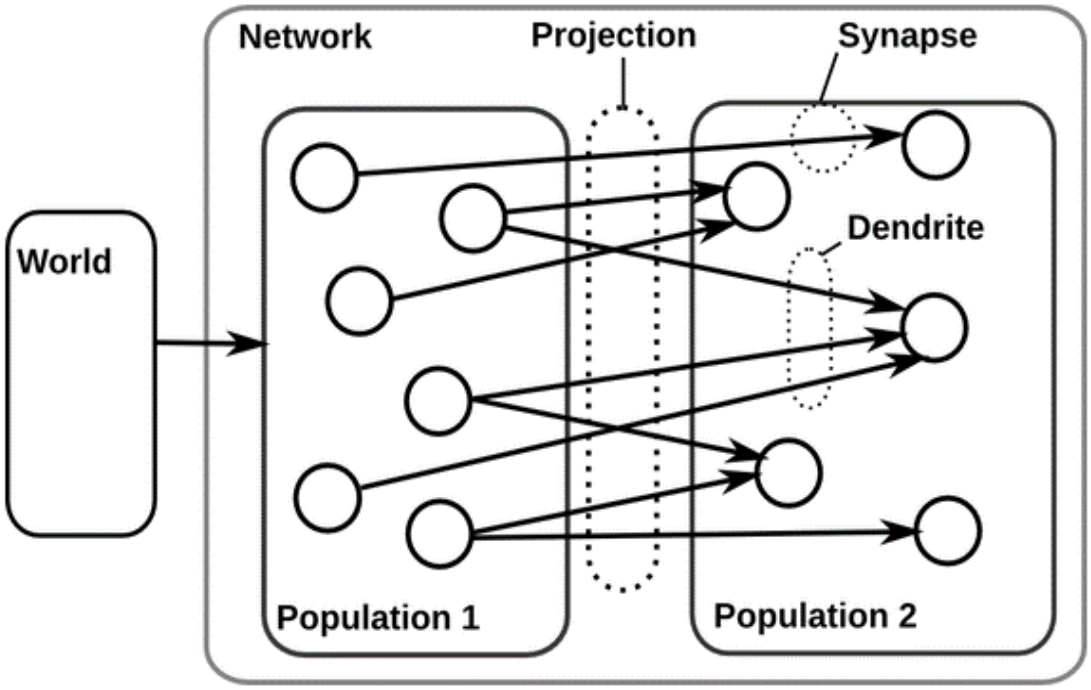


Figure 5: Neural Networks [15]

The neural networks are consideredas the most common network in the network architecture. The units in the neural network also provide the sum which is related to the inputs, and after this, it passes the activation to the transfer unit so that the output function is given in the neural network. The networks provide the simple interpretation in the form of input and the output model [7]. The neural networks show the function of the uninformed complexity with the number of layers. It also provides the number of units in each layer which determines the functional complexity of the networks. The issues which occur in the multilayer perceptron include the condition of the number of hidden layers and the number of units in the network. The number of input and the output unit is defined by the problem which occurs in the neural network [10].

The inputs variable in the neural network shows the hidden layers. The number of hidden units are far from each other so that they perform the functions in an equivalent manner. The hidden layers consist of the units. Half of the units are related to the input units, and the other half are related to the output units in the neural network [2].

1.2.1.3. Linear Regression

Linear regression is used in the machine learning to predict the output data with the use of new data based on the previous data. Linear regression is one of the best models which help to predict the hidden parameters. It is mainly based on the generalized linear model. In the linear model, the output variable is assumed to be taken as the linear combination of input variables [10].The output variables in the linear regression take the continuous values. The output and the input values in the linear regression are taken in the numeric form [11].

1.2.1.4. Random forest

Random forest technique is used in machine learning which utilizes the combination of the bagging. It also takes the random selection of the process either in the regression form or classifies the problems into the different units. The one motivational work of this algorithm is based on the insensitivity of the variance and the results which are predicted from the units to measure the price value of the project[10].

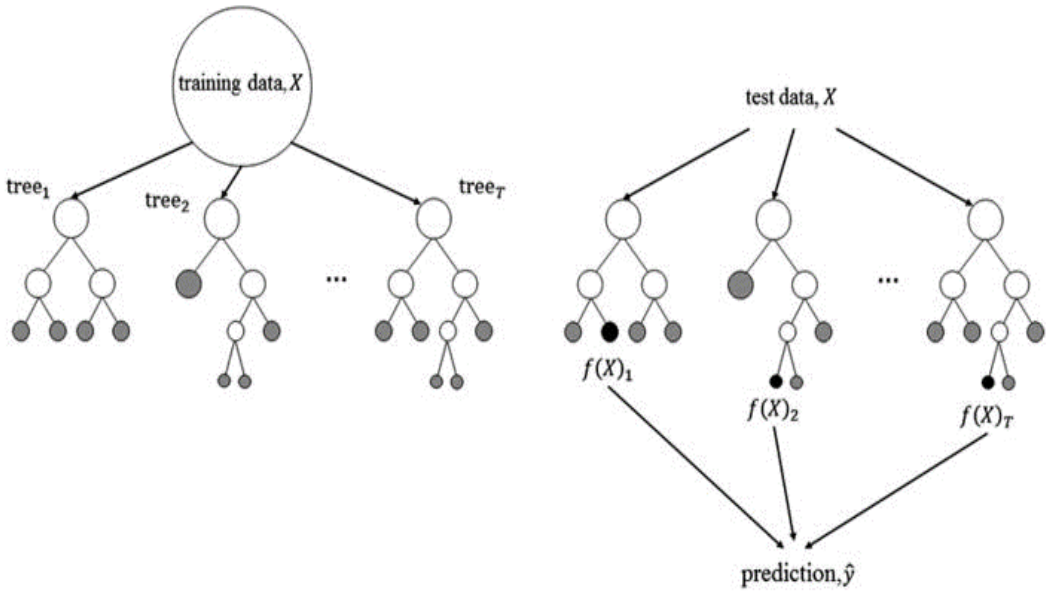


Figure 6: Conceptual diagram of Random Forest [51]

From the other machine learning techniques, Random Forest is the best technique which provides the better results from the units. It is because of its nature of randomly selecting features which provide the aggregating mechanism from the given outputs. The above figure shows the conceptual diagram of Random Forest algorithm. Firstly, the trees are trained individually through recursive binary partitioning of sample data. Then the test data is dropped down to every tree, and the final response is an average of all predictions in the forest.

Another benefit of the random forest is that it takes only a few parameters to set the implementation of the units. For example- take the 250 movies from the imdb.com and predicts the ratings of the movie based on their features. Then the following approach will be used-

1. Load the base data from the imdb.com
2. Now utilize the gross revenue data from the imdb.com website and then merge that data into the base data frame.
3. Now clean the data by using the followings tasks- change the released data type into the daytime and changed the runtime unit into the minutes. The last step is to change the year into the integer.
4. Now change the entries into the null values. After this, change the IMDB data into the float. Now remove the comma from the numbers and the variables.
5. Now create adummy variable in the data and merge them into the base dataframe.
6. Now take the top ten movies based on the gross revenue and take the top 10 actors from the 250 movies.
7. Now create a simple machine learning model which uses the Random forest technique based on the five features.
8. Now use a boosting algorithmto compare the results and to improve the accuracy.

1.2.1.5. Decision tree

The decision tree has many equivalenciesin real life, so it influences the wide area of the machine learning. The decision tree covers the both units: the classification units and the regression units. In the decision analysis, the decision tree can be used

for visual and the effective results which represent the necessary decision, and provides us the techniques for decision making [3].

The decision tree uses the tree model for taking the decisions in the units. Itis commonly used in the data mining which covers the strategies to reach their particular goal. The decision tree is also widely used in the machine learning. The decision tree algorithm first takesdata and then a decision tree is drawn with its root at the top. After this, the tree splits into the branches and the edges. The end of the branch which does not split shows the decision. The leading feature of this algorithm is that it offersthe clear and the visual results [10]. The figure below shows the basic structure of Decision tree.

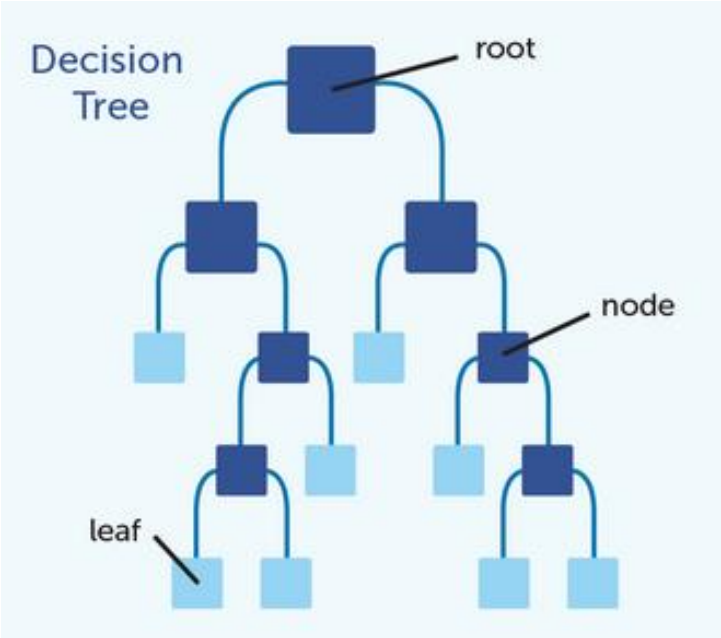


Figure 7*:* Decision tree structure [53]

1.2.1.6. Adaptive boosting

AdaBoost is also known as the adaptive boosting which uses the multiple units to generate the strong composite learner. AdaBoost is a machine learning algorithm which is used for the classification and the regression analysis. AdaBoost is the most resistant technique which is used in the machine learning. It takes the sensitive and the noisy data from the input and the output units.AdaBoost creates the strong learner in the units by adding the weak learners in the technique [11]. This process is

applicable in each round of training. The new weak learner also added in the algorithm in the form of the vector unit which adjusted to focus on the examples in the previous units. The result is based on the high accuracy as compared to the learners in the units.

1.2.1.7. Naive Bayes

In machine learning, we use best methods and the techniques to find out the best results from the input and the output units by taking the appropriate figures in the machine learning techniques. The naïve Bayes theorem is used to find the hypothesis results in the given data. With the use of the naive Bayes theorem, we can select a possible hypothesis from the given data so that we can use the preceding knowledge about the problem [4]. The naïve Bayes theorem provides the ways so that the probability can find out from the preceding knowledge.

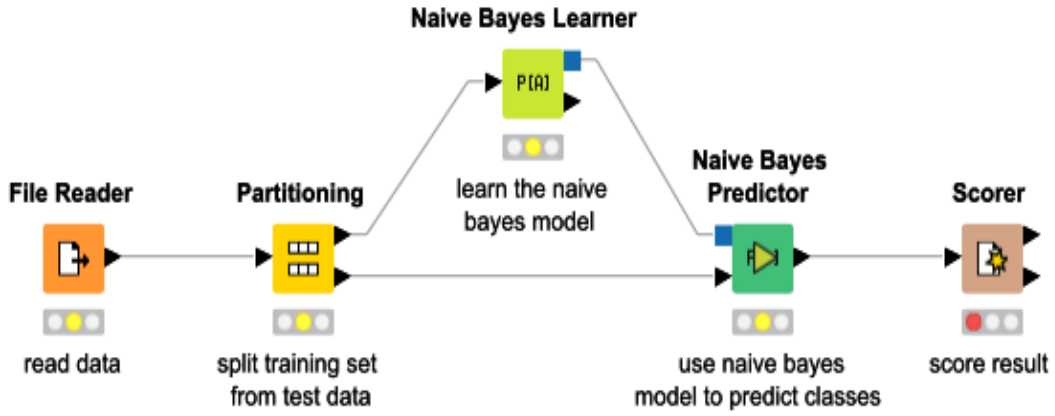


Figure 8*:*Naïve Bayes algorithm process

Bayes theorem started from the following given formula-

$$P(h|d) = (P(d|h) * P(h))/ P(d)$$

Where,

- P(h|d) is the possibility of the hypothesis h, and the d considered as the given data. It is also known as the subsequent possibility.

- P (d|h) is the possibility of the data d which proved one thing that the hypothesis h was true.
- P(h) is the possibility of hypothesis, and this is also called the prior possibility.
- P(d) is the given data.

In the Naïve Bayes theory, the hypothesis possibility and the given data is to be used while calculating the results in the given data.

1.2.1.8. Binary Discriminant

Binary Discriminant is an algorithm which is used in the classification problems of the given data. If there are more than two classes, then the binary discriminant analysis will be preferred to the linear classification technique. The binary discriminant analysis involves the predictive classification models which give the results from the output and the input units.

The binary discriminant model is used for both the preparation and the application units which prepared the data to find the results in the figures so that the accurate result can be obtained for the further proceeding. The binary discriminant consists of the collected data which provides the properties to the data and then helps to calculate the data in each class [2]. The Binary Discriminant use a single input variable and consider that variable as the mean variable in the data. But for multiple variables, it uses the same properties which calculate the multivariate Gaussian and the covariance matrix.

How to prepare the data for the binary discriminant analysis?

- Classification problems- The binary discriminant used for the classification of the problems where the input and the output data will be categorized according to their attributes.
- Gaussian Distribution- The binary discriminant model also used the Gaussian distribution from the input variables. The Gaussian distribution has different attributes which involve the exact variables in the units.
- Remove Outliers- The remove outliers used the basic statistics which used to separate the data in the binary discriminant. It separates the data into the mean and the standard deviation.

1.2.1.9. Boosted tree

Boosted tree is also known as the gradient boosting. The gradient boosting is mainly based on the original models in the machine learning. It is an implementation which involves the various tools to find the results from the units. It belongs to the broader collection of the data under the distributed machine learning community. Gradient boosting is mainly applied to the machines and access many functions in the machine.

- It also provides the command line interface.
- C++ language in the machines.
- It also creates the Python interface in the model.
- It also provides the model in the caret package.
- Enables the JAVA and the JVM languages in the machines so that they perform various functions.

Algorithm features of the Gradient are boosting:

By enabling the gradient boosting, one can save time and memory resources. The main goal of this algorithm is to make the best use of the available resources and to take benefits from that resources [11]. The key features of the algorithm are-

- Sparse aware:It implements the sparse aware to handle the automatic data or to find the missing values in the data.
- Block Structure:It supports the decision tree and makes them parallel to each other by constructing the tree in the structure.
- Continued training: It helps to boost the fitted model in the new data.

1.3. Ensemble Models

Ensemble modeling is the technique of running two or more different analytical models to synthesize the result in a single score so that accuracy of the predictive analysis and applications of data mining can be improved [25].In the process of ensemble modeling, prediction group of base models is combined to generate the

composite predictions which give more accuracy. There are two main activities of ensemble modeling which are given below:

1. Constructing base learner ensemble from the training data
2. Combining predictions of ensemble into composite prediction

Many types of the base learners are used in the ensemble learning, and there are various ways to generate the composite predictions. But combinations of all ensemble constructions and base models are not always useful [34].The main objective of the ensemble model is to construct the high accurate predictive model.

Ensemble models are a learning technique which combines the various weak learners or models so that they produce the strong model which works better. The benefits of ensemble models are given below:

- Better prediction
- Give more stable or constant model
- Better Forecasting
- Better Results
- Decrease the Errors

An example of the ensemble model is a random forest model which is a data mining leverage multiple decision trees used to predict the outcomes based on various rules and variables. A random forest model analyses the sample data and then evaluate the various factors or variables differently [47].

1.3.1. Methods of Ensemble Models

Ensemble methods use the various algorithms in machine learning to get the better predictive performance. The different ensemble methods are given below:

1.3.1.1. Bagging

Bagging is used for the bootstrap aggregating. It is simplest ensemble based algorithm which provides better performance. In this type of algorithm, each model has an equal weight in the ensemble vote. In this method, firstly random samples of

the training data set are created, and then a classifier for each of the sample builds. At last, the results of this different classifierare combined with the help of majority or average voting. This type of method is used to decrease the variance error [18].

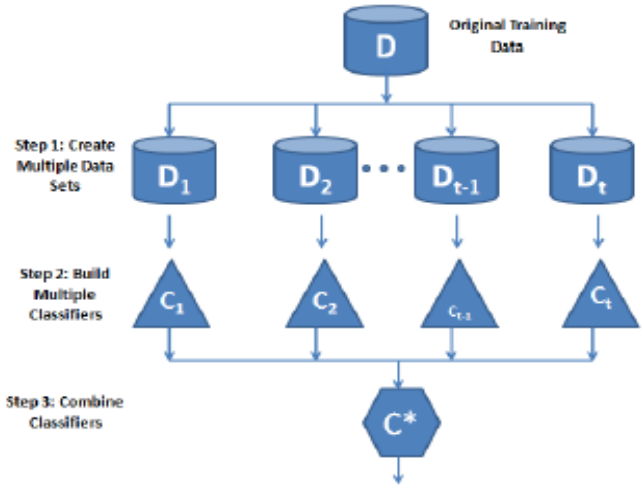


Figure 9: Bagging

1.3.1.2. Boosting

It is another ensemble method which provides sequential learning of the predictors. It refers to the group of an algorithm which converts the weak learner to the strong learner. It also generates the classifiers ensembles on the various data which are integrated by the majority of the votes. In this method, each of the training set has some weight, and each set weight is updated on the iteration. In comparison to the bagging, boosting method provides better accuracy and result.But it has a limitation as it is not suitable for the training data [20].

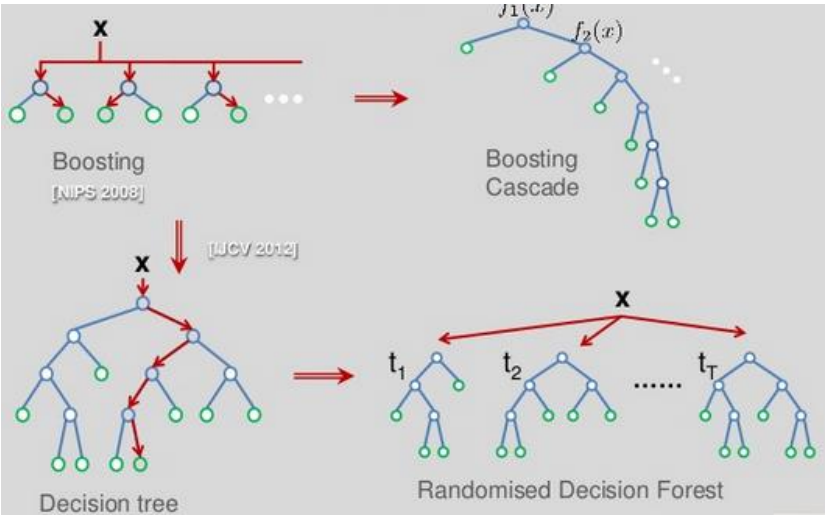


Figure 10:Boosting process [52]

1.3.1.3. Stacking

It is also known as the stacked generalization used for unsupervised learning and supervised learning**.** It is used to combine the predictions of another machine learning algorithm**.** In this method, firstly all other algorithm is trained on the data set to combine the algorithm used for the final prediction**.** This algorithm performance is better as compared to the individual trained model**.** It works on two phases i**.**e**.** use the multiple base classifier in the first phase to forecast the class and then in next phase new learner is used to joining their predictions which reduce the error**.**

1.3.1.4. AdaBoost

It is the machine learning method where various learners are used to constructing strong learning algorithm that works on the principle of thebase algorithm and iteratively is used for improving it for the classified training dataset**.** In this type of algorithm, equal weight is assigned to the training data to select the base algorithm**.** In each, the stage base algorithm is used in the data set to increase the weight of incorrect classified data**.** It decreases the error during learning**.**

1.3.1.5. Bucket of Models

In this method, the model selection algorithm is used to select the optimum model for each of the problem**.** If this model is tested on one model, then it does not give the better results**.** The best method used in model selection is cross validation selection which analyses and compare the algorithm of machine learning by dividing the segments**.** One of the segment is used to train the model, and another one is for validating the model**.**

Chapter 2

Literature Review

2.1. Analytical Approaches

B.U. Nwosu et al. [1**]**described the status of the Vitamin D in pediatric irritable bowel syndrome**.** Irritable Bowel Syndrome (IBM) disease is mainly found in the children and adolescents whose treatment options are very limited. IBS due to the deficiency of Vitamin D can cause migration, anxiety, and the depression. It described the pediatric patients suffering from IBS have lower 25(OH)D concentration which is similar to the BMI values as the controls. The result concluded that the 7% of the children and adolescents are suffering from IBS have Vitamin D sufficiency, and less than 50% of the children and adolescents have Vitamin D deficiency. The prevalence of the Vitamin D is much greater than IBD and malabsorption syndromes. IBS has the high prevalence of Vitamin D due to the lifestyle habits which limit the exposure to the sunshine, restricted food choices, and hypoalbuminemia. The close monitoring of the status of Vitamin D is checked by the routine clinical care of the patients who are suffering from the IBS. Many control trials identified the effect of the supplements of the Vitamin D on extra-intestinal manifestations and intestinal of IBS.

B. Wingfield et al. [2**]** presented a novel combination of Multilayer Preceptors **(**MLAs**)** and Support Vector Machines (SVMs**)** in a conditional multiple classifier symbols which describe the subtype, presence, and the activity of the IBD from a stool stable that avoids the needs of the colonoscopy**.** Multilayer Preceptors **(**MLAs**)** and Support Vector Machines (SVMs**)** are used with the datasets of microbiomes**.** An experiment was performed in this research, using computational intelligence algorithm and machine learning to determine that the new features and the alternative algorithms are appropriate or not**.** MLAs and SVMs both showed the better and good performance when they were applied to the classification of the metagenomics**.** Human Microbiomes are used to analyze the bacterial group which is present in the gut**.**

Inflammatory Bowel Disease is an inflammatory disease caused in human colon and the small intestine. The symptoms of the IBD are non-specific whose diagnosis are delayed due to an invasive colonoscopy which is required for the confirmation. Poor health among the children is due to the delayed diagnosis. In this paper, the analysis was conducted for the microbiome using shotgun sequencing**.** Predicted metagenome was produced from the 16S marker gene survey with the help of PiCRUSt which was validated by the Boruta algorithm**.**The functions of the genes are identified by reconstruction of Unobserved states with the help of philological investigation of communities which are used to represent the new features of the source**.** For the classification of the disease, functions, and features of metagenomic are used**.** The features which are identifiedwere implemented in the pathogenesis of IBD and IBD relapse**.** The features were investigated using the Boruta algorithm which concluded that many of the features are from the metagenome.

T.G. Dietterich [4**]** explained about the ensemble methods in the machine learning**.** Ensemble methods are the learning algorithm used in the machine learning to construct the set of classifiers and new data points by the predictions vote**.** The method which that was used in this research is Bayesian averaging but therecent algorithm used error-correcting output coding, boosting, and bagging**.** In this research, these methods are described and also explained the reason for the better performance of ensembles as compared to the single classifier**.** Ensembles used to obtain the highly accurate classifier which is obtained by combining the less accurate**.** In this research, the methods used for constructing the ensembles are Bayesian Voting, Manipulating the training, manipulating the input and output target and Injecting Randomness**.** Bayesian Voting used the following equation to construct the ensembles**:**

$$P(f(X) = y|S, x) = \sum_{h \in \mathcal{H}} h(x)P(h|S)$$

From the results, it is concluded that the performance of the ADABOOST is good and performs better in the ensembles**.** The machine learning algorithm is combined with the ADABOOST which concluded that the algorithm is of global character**.**

E. Consgunet al. [5**]** proposed a high-dimension pharmacogenetics prediction of the continuous trait with the help of machine learning techniques which has the

application of warfarin**.** Complex traits and diseases contribute to many of the genetic factors while the current genotyping arrays include a single technique which is needed to the comprehensively model of the genetic variance**.** There are many benefits of using machine learning techniques such as high flexibility, power, and lack of parametric assumptions**.**

For the prediction of warfarin maintenance dose in a cohort of African Americans, three methods were used i**.**e**.** Support Vector Regression **(**SVR**)**, Boosted Regression Tree **(**BRT**)** and Random Forest Regression **(**RFG**).** Amulti-step method is developed which is used to build prediction models and selects the SNPs with the help of various subsets of selecting SNPs, genetic which are associated with them, and the environmental variables**.** These discovered models are tested in a cross-validation framework**.** From the results, it is concluded that the modelling approach provides higher accuracy as compared to warfarin dose prediction and a model which has size 200 SNPs provides the higher and best accuracy**.** In the warfarin dose, the R2 is between the predicted and actual square root was 66**.**4**%** for the RFE, 56**.**9**%** for BRT and 57.8 **%** for SVR**.** Among these RFG has the better and high accuracy**.**

K.I. Penny et al. [6**]** observed the health-related quality of life in a cohort of persons with irritable bowel IBS and determined the socio-demographic with the help of data mining methods**.** In this paper, it is concluded that the symptoms of irritable bowel syndrome are associated with the quality of life-related to health which is affected badly by IBS**.** The predictive factors influence the quality of life of the IBS patients**.** A cross-sectional survey design is developed for this in which the general population of UK is considered**.** The quality of life-related to health was determined with the help of a battery of validated questionnaires**.**

Datamining models are used to determine the factors of the impaired quality of life-related to health, artificial neural network, classification tree, and logistic regression**.** From the results, it is concluded that sociodemographic and psychological morbidity influence the health-related quality of life of a patient with irritable bowel syndrome**.** Health-related quality of life of individuals with irritable bowel syndrome is impaired in the UK**.** This paper also confirms the impact of the irritable bowel syndrome on HROoL in the community-based patients suffering from IBS which is underestimated**.** Data mining technique is poor predictive on HROoL in irritable

bowel syndrome**.** Data mining method is also referred as the black box method due to less intervention**.**

Nidhi et al. [7**]** explained about the prediction of biological targets for the compounds with the help of Multiple-Category Bayesian Models on the databases of Chemogenomic**.** In this paper, firstly the target identification for the small molecules are identified which elicit the biological phenotype which provides a silico correlate of target fishing technologies for outing the fish rapidly for the compounds which are thebasis of chemical structure**.** A Laplacian-modified naïve Bayesian model has been developedin the WOMBAT Chemogenomic databases **(**World of Molecular Bioactivity) on the extended connectivity fingerprints of compounds. This model is developed to identify the protein target for the MDDR (MDLDrug Database Report) database compounds. From the result, it is foundthat 77% of the time is for the compound from 10 MDDR activity classes that are associated with the genetic activities.

I.Kurt et al. [10**]** compared the performance of the classification techniques, so that presence of coronary artery disease can be predicted **(**CAD**)** for which a retrospective analysis is performed in 1245 subjects in which 865 has the presence of CAD and 380 has an absence of CAD**.** In this paper, performances are comparedto the classification and regression tree **(**CART**)**, Radial basis functions **(**RBF**)**, Multi-layer perception **(**MLP**)**, logistic regression **(**LR**)**, and self-organizing feature maps **(**SOFM**).** Some of the predictor variables are sex, smoking status, age, diabetes mellitus, hypercholesterolemia, systematic hypertension, and body mass index **(**BMI**).**

ROC curve, Multidimensional Scaling **(**MDS**)**, and Hierarchical Cluster Analysis **(**HCA**)** are used to compare the performances of the classification of the techniques**.** The areas for MLA, LR, CART, RBF, and SOFM which comes under the ROC curves are 0**.**783, 0**.**753, 0**.**745, 0**.**721, and 0**.**675 respectively**.** The best technique which is used for the prediction of the presence of the CAD in the data set is MLA which gives the better classificatory performance**.** From the result, it is concluded that according to MDS and HCA, MLA, LR, RBF, and CART performed better as compared to the SOFM in the predicting CAD**.**

M.A. Razi et al. [11**]** performs the comparison of prediction accuracy which includes models of CART, NNs, and nonlinear regression with the help of continuous

dependent variable, categorical predictor variable and a set of dichotomous**.** On the smokers, a large dataset is used to run these types of models**.** The errors in the prediction are compared in the dependent variables which are continuous and in predictor variables these are categorical**.** CART and NNs model produce better prediction accuracy as compared to the non-linear regression model**.** In this paper, they described that NNs model produces the lower value of MSE, MAPE, and MAE as compared to the CART model**.** From the result, it is also concluded that the decision tree based models are scalable to the large problems and can handle the smaller data than the NNs models**.**

BothNNs and CART model provide a satisfactory prediction, but the regression model is slightly better than these models in terms of performance in model verification and construction**.** The performance of the NNs model is better than the CART model on the multimode classification problems where the data set is very large containing few attributes**.** The data sets present in the CART outperform NNs models are smaller which has a large number of irrelevant attributes**.**

R. A. Awad et al. [12**]** describes the defaecography in the patients who are suffering from the irritable bowel syndrome**.** Many of the symptoms are observed in the origin of the recto-anal segment with the internal anal sphincter, alternations in the rectal sensitivity and the motility changes in the rectum in the patients which are suffering from the IBS**.** In this paper, 16 patients with the IBS and 10 of the healthy volunteers are evaluated with defaecography, and three results are found**:**

- **Anorectal angle:** There is no significant difference seen in the anorectal angle but during defaecation and rest between the patients with IBS and healthy volunteers**.** During the defaecation, patients with IBS were unable to widen the angle**.** The patients with IBS constipation showed no difference at rest and during defaecation as compared to normal frequency defaecation**.** During the defaecation, the healthy volunteers widened the angle more than 5°.
- **Perineometer**: During the simulated IBS patients had less perineal descent defaecation as compared to the healthy volunteers**.** Also, there is less mobility or perineal descent in the IBS patients during the squeeze**.**

From the result, it is concluded that the IBS patients with constipation predominant or not shows the changes in the mobility of pelvic-floor**.**

J.K.Zia, et al. [17] confirmed the feasibility and usability of the traditional paper food and the GI *(*Gastrointestinal*)* symptom journals for the patients which are suffering from IBS**.** The Same type of results is observed for electronic version journal**.** So, the result concludes that the data which is collected either from paper or electronic journal are accurate and reliable**.** None of them demonstrate the significant impacts on the GI symptoms**.** More than the half number of participants found that journaling is clinically useful and they want more feedback and the guidance about their dietary**.** The analysis done in this paper described the relationships between the GI symptoms and the meal nutrients**.** They described that the paper food and the GI symptoms are useful for the IBS patients who determine the potential trigger food**.** This paper evaluates feasibility, usability and clinical usability of the journals as data collection tool**.** And then explore the method for analyzing the journal data so that they can describe the symptoms and diet patterns**.**

R. L. Soares [18]described the Irritable bowel syndrome (IBS) which is a clinical challenge in the today's 21st century. It is the most diagnosed gastrointestinal condition. IBS can affect up to one people among the five people at the same point in their lives which significant impact on their health care utilization and the life quality. The biopsychosocial model is used for the prevention for IBS among all the proposed theories and mechanisms. The complex of the symptoms would lead to the interaction between psychosocial, behavioural, psychological, and the environmental factors. With the structural abnormality and specific test, the diagnosis of IBS cannot be confirmed. Today, we use the current gold standard Rome Criteria III for the diagnoses of the IBS. For the purpose to diagnose IBS, there is no clinical evidence which recommended the use of biomarkers in the blood. There is no definitive treatment for the IBS, but it can be controlled by non-pharmacologic management eliminating by some exacerbating factors such as stress condition, drugs, and changes in dietary habits. The traditional pharmacologic management of IBS used the several drugs, and it is based on the symptom.

O. Grundmann and S. L. Yoon [19]presented the alternative and complementary medicines in the Irritable Bowel Syndrome. It is the common gastrointestinal disorder in the general population. The diagnosis of Irritable Bowel Syndrome is based on the intestinal conditions in which specific antiques and inflammatory markers are absent. The current methods of the pharmacological treatment concentrate on minimizing the

symptom severity which results in effectiveness gap for the IBS patients in increasing their quality of life. In the IBS patients, complementary and alternative medicines (CAM) are also included in a higher degree of symptom management and quality of life. It is concluded from the past decades that a number of important clinical trials have specific herbal therapies, cognitive behaviour therapy, yoga, hypnotherapy is present which give the improved outcomes of the treatment in the IBS patients. They also propose an integrative method to treat the diverse symptoms of the IBS by combining the need and benefits for the pharmacotherapy so that they can provide the best treatment to the IBS patients.

Q.L. Tang, et al. [20] demonstrated the cognitive-behavioural therapy for the management of the irritable bowel syndrome. IBS is the common type of the disorder and chronic relapsing condition which is associated with the specific disability. It also has a financial burden for the services of health as many of the resources are consumed such as the cost of treatment, investigation, and physician time. The symptoms of psychiatric in the IBS patients is due to the psychotherapy or CBT. CBT has implications of his own clinical practice is for gastroenterologists. There ismuch evidence of psychological and physical symptoms of the efficacy of CBT of IBS patients. CBT is the bestrecommended treatment which is provided to the IBS patients, but it has some psychological distress and physical discomfort.

T.A. Majid, et al. [21]presented the quantitative analysis on the level of acceptance of IBS in the Malaysia construction industry**.** The study on the IBS usage is carried out from June 2008 to December 2008**.** For the acceptance of the IBS, it is measured using the TAM which includes a***)*** perception on the usefulness of IBS, b) the awareness of the IBS system, c) the perception of the ease of IBS usage and d) the actual usage of IBS. The average percentage difference of the t-test shows that the respondents have agreement on the use of IBS that the problems which are faced in using the IBS and the awareness on IBS but there are also some disagreements which show that the contractors also have common problems in using the IBS.

In the construction industry, there is much importance of the IBS and also need to overcome all the problems related to IBS using. In the Malaysia construction industry, CIDB is used so that they can promote the IBS. The t-test concludes that there has asignificantdifference in the variable except for their actual use of the IBS. CIDB has

the important role in educating the construction industry for applying the IBS extensively. This paper shows that according to the contractors, IBS system can help in providing the better-quality building systems and also minimize the dependency on the foreign workers.

V.Sankar, et al. [23]presented the simultaneous fecal microbial and metabolite profiling which enables the accurate classification of the pediatric irritable bowel syndrome**.** They also describe the fecal microbiota and metabolite differences between these two adolescent populations which used for discriminating the difference between the health and IBS**.** For this, they constructed the individual**'**s classification models for microbiota **–** and metabolite sample which is based on partial least squares multivariate analysis and then applied them to the Bayesian approach so that they can integrate the individual model into the single classifier**.** The combined classification results give the 84*%* accuracy of correct sample group assignment and 86*%* for the IBS-D in the test of cross-validation**.** The performance of the cumulative classification model is validated with the help of the de novo analysis of the stool samples from IBS-D cohort**.** From the result, it is concluded that the high-throughput metabolite and microbial profiling of stool sample can be used to facilitate the diagnosis of the IBS**.**

E. Giannetti and A. Staiano[26]summarize the evidence related to the probiotics treatment for the pediatric IBS**.** They find that the management of the children suffering from IBS should be tailored to the patient who has specific identifiable triggers and symptoms**.** The main therapeutic method is complementary*/*alternative medicine, pharmacologic, psychosocial, and dietary interventions**.** The evidence for the pharmacological therapies is very less like antidiarrheals and antispasmodics**.** The review of Cochrane concluded the beneficial effects of the pharmacological agents which provide the relief from the functional abdominal pain symptoms in the children**.** The role of antibiotics is controversial for the treatment of the children suffering from IBS**.** There are also many non-pharmacologic treatments are used for the pediatric IBS**.** Some of the beneficial effects found from the evidence are hypnotherapy, cognitive behavioural therapy, hydrolyzed guar gum, and probiotics**.** Few of the randomized clinical trials are present in the children**.**

For the treatment of the Functional Gastrointestinal Disorder, a meta-analysis of 9 trials tested with different probiotics in the adolescents and children. From this, it is concluded that Lactobacillus reuteriDSM 17938, Lactobacillus GG and VSL#3 are very helpful in the treatment of the IBS. A mixture of the longumBB5361, BifidobacteriuminfantisM-631, and breveM-16V1 is safe in children suffering from irritable bowel syndrome and also have the better control on AP, and improved the quality of life as compared to placebo. Due to the importance of the gut microbiota in influencing brain-gut interactions, probiotics are used as the therapeutic tool in FGIDs.

S.H. Park, et al. [28]investigated the evidence for using the relaxation therapies as the interventions which reduce the symptoms of the irritable bowel syndrome *(*IBS*)* and its severity. They also decrease the anxiety and helps in improving the quality of life of the patients which are suffering from the irritable bowel syndrome. For this, the electronic bibliographic database was used which identify the randomized controlled trials that include in the programs of the relaxation exercise for the adults with IBS. In this paper, Cochrane's risk of bias is used to access the study quality. It is concluded from the result that the IBS symptoms are significantly decreased due to the use of the Cochrane's risk of bias, and there was no heterogeneity.

The results of the systematic review are based on the eight RCT which conclude that the relaxation techniques have the positive effects on the physiologic symptoms of patients suffering from IBS. There was no difference in thequality of life and anxiety among the various groups. So, there is need of the careful interpretations of the results. Under the different relaxation techniques, muscle relaxation is the main technique. From the result, it was concluded that the relaxation methods have effective and accessible interventions for the nurses to use with the patients suffering from IBS. There is also a need for the long-term research which reduces the sociologic, physiologic and mental problems in the patients suffering from irritable bowel syndrome.

M.Mustafa, et al.[30]described the pathophysiology, management and the treatment of the irritable bowel syndrome. Irritable bowel syndrome has the high prevalence in Brazil, Mexico, and Pakistan and low prevalence in Canada. The

etiology of the irritable bowel syndrome is unknown for the brain-gut axis, post-infection, small intestine bacterial overgrowth, high stress, genetic defects relating to the immune system, anxiety levels, and protozoal infection are some of the causes of the patients which are suffering from the irritable bowel syndrome**.**

For the deployment of the irritable bowel syndrome, psychological and genetic environment is important**.** The diagnosis algorithm which is used for the irritable bowel syndrome is obsolete Rome 1 and 11 criteria, Manning criteria, and Kuriscriteria, Rome 111 process. The treatment of the irritable bowel syndrome is by mesalazine, antispasmodic, aminosalicylate, antidepressant, and the stress management**.** The beneficial effects of probiotics on the IBS symptoms is by treating small intestinal bacterial growth, decreasing small intestine permeability, gut microbiota, improving intestinal transit time, and normalization of cytokine blood levels. The symptoms of the irritable bowel syndrome are discomfort, abdominal pain, constipation, and diarrhea. Irritable bowel syndrome is referred as the functional disorder which has no organic cause**.** Misdiagnosed Irritable bowel syndrome leads to ahigh impact on the health of patient which also includes the consequences of socioeconomic**.**

H. **Vahedi, et al**. **[31]**described the Irritable bowel syndrome which is the most prevalent functional gastrointestinal disorder in the general population**.** Its signs, chronic nature, and the symptoms vary periodically from mild to severe, and they also have the negative effects on the quality of life for the IBS sufferer**.** So, there is need of the appropriate treatment for the IBS patients**.** Patients should be informed about the type of diseases by their doctors is benign, and they should also be educated to deal with the disease and the ways to control the symptoms of the Irritable bowel syndrome diseases**.**

M. E. Mcomber and R. J. Shulman [32]demonstrate about the recurrent abdominal pain and irritable bowel syndrome in the children**.** Recurrent abdominal pain is one of the most ubiquitous conditions which was faced by the healthcare team, and it also has the economic and emotional impacts**.** The psychological condition is used for recognizing the environmental and physiological contributions, and it is used to consider the condition of the framework of the biopsychosocial model in which social, biology and psychology environment interact**.** They describe the diagnostic,

etiologies methods and the treatments methods for recurrent abdominal pain in the children**.** The psychological state of child and the parent in terms of somatization, anxiety, and coping skills modulate the expressions of the symptoms**.** The newer treatment methods are distraction and relaxation therapies and the medications**.** The main reason for the morbidity and emotional distress in the pediatrics is due to the functional gastrointestinal**.**

Y. Tanaka [37]described the biopsychosocial model of Irritable Bowel Syndrome which is a chronic disorder in the Gastroenterology**.** It is characterized by the discomfort and the recurrent abdominal pain with distributed bowel function**.** Basically, it is a heterogeneous disorder whose treatment is varying, and the physicians struggle in order to find the optimal treatment for the patients with the Irritable Bowel Syndrome**.** It includes the high health care cost, and it also decreases the health-related quality of life**.** It results from the enteric and central nervous system interactions**.** The factors which are related to the psychosocial are cognitions, physiology, illness behaviour and manifestations**.**

It is essential for the physician to find the issues of the psychosocial of the patients which are suffering from Irritable Bowel Syndrome and create a good relationship between physician and patient in order to optimize treatment of the disease**.** They described the a***)*** the psychological variables related to life stress, the symptom performance, and their related patterns, b) the predisposing psychological variable are seen in early life, c) the clinical results and effective treatments comprising psychotherapeutic techniques and d) gut pathophysiology with focussing disturbances in motility, visceral hypersensitivity and brain-gut interactions which also include the methods of psychotherapeutic.

Table 1: Summarization of various analytical approaches

Authors	**Year**	**Objective**	**Method**	**Result**
B.U. Nwosu, L. Maranda,	2017	Characterize the vitamin D status of patients with IBS	25-hydroxyvitamin D [25(OH)D] concentration	IBS had significantly lower 25(OH)D 7%children and

N. Candela				adolescents with IBS were vitamin D sufficient >50% of the subjects with IBS had vitamin Ddeficiency.
B. Wingfield, S. Coleman, T.M. McGinnity, A.J. Bjourson	2016	Investigated feature relevance of IBM	Boruta Algorithm	MLAs and SVMs gave better performance Majority of relevant features were taken from the predicted metagenome
T.G. Dietterich	2000	Reviews the ensembles methods, Reason for the better performance of the ensembles	Bayesian averaging	Performance of the ADABOOST is good and performs so well in the ensembles
E. Consgun,N. A. Limdi, and C.W. Duarte,	2011	Prediction of warfarin maintenance dose in a cohort of African Americans	Random Forest Regression, Boosted Regression Tree, Support Vector Regression	66.4% for RFR, 57.8% for SVR and 56.9% for BRT. Thus, RFR had the best accuracy.
K. I. Penny and G. D. Smith	2009	Evaluate the health-related quality of life in a cohort of individuals with irritable bowel	Data Mining	Sociodemographic and psychological morbidity influence the health-related

		syndrome		quality in IBS**.**
Nidhi, M. Glick, J. W. Davies, and J. L. Jenkins	2006	Prediction of biological targets for the compounds	Multiple-Category Models	Improves the knowledge in chemogenomic databases
I. Kurt, M. Ture, A. T. Kurum	2008	Predicting Coronary Artery Disease	Comparing performances of logistic regression, classification and regression tree, and neural networks	MLA gives abetter classificatory performance for the presence of CAD**.**
M. A. Razi, K. Athappilly	2005	Comparison of Prediction accuracy involving nonlinear regression, NNs, and CART models	Continuous Dependent Variable and a set of Dichotomous and Categorical Predictor Variables	NNs and CARTmodels provide better Prediction than the regression models

R. A. Awad, J. Martin, M. Guevara, R. Ramos, J. L. Noguera, S. Camacho, R. Santiago, J. L. Ramirez, A. Toriz	1997	Defaecography in the patients which are suffering from the Irritable Bowel Syndrome	16 patients with the IBS and 10 of the healthy volunteers are evaluated with defaecography	No significant difference was seen in the anorectal angle IBS patients had less perineal descent defaecation than healthy volunteers

2.2. Evolutionary Approaches

A.S.M. Salih et al. [3**]** evaluated the Ensembles design and combined different algorithms which were used to develop the Novel Intelligent Ensemble Health Care Decision Support and Monitoring System**.** This proposed systemwas used to classify the emergency hospital situation on vital signs from wearable sensors**.** For this purpose, the number of attributes was observed to be decreased from 300 to 6**.** In the process of monitoring the healthcare, data mining technique was used to predict and classify the diseases**.** The main purpose of this paper was to construct the new Novel Intelligent Ensemble Health Care Decision Support for the intelligent health monitoring system and also reduce the dimensionality of the attributes**.** The experiment was performed on the wearable sensors which uses the environment of the hospital**.**

- Firstly, the performance was classified using various classifiers**.**
- Then the performance of different Meta base classifiers was compared using Voting, Stacking, Random Committee, Logic Boost, AdaBoostM1**.**
- Investigated the Meta classifiers and new Novel Intelligent Ensemble method

In this research, different ensemble combining models are explored and evaluated using different methods which are based on the ROC curves, Error Metrics, Specialty, Confusion Matrix, and the Cost/Benefit Methods. Then the performance of classifiers was compared, and from the results, it was concluded that Voting combining with the J48, Random Tree, and Random Forest, gives the high recall, high f-measure, and better accuracy. The Novel Intelligent Ensemble Health Care Decision Support and Monitoring gave the optimized results and improved the health care monitoring.

N. Arsov et al. [8**]** observed the highly accurate prediction of hypotheses with the help of the collaboration of the ensemble learning. Ensemble generation is the convenient method of achieving the best performance of generalization for learning the algorithm by collecting the predictive capabilities. For the binary classification of the ensemble, bagging and boosting are combined which improves the stability through variance reduction. A multi-model is used to combine strives for net-balancing the bias-variance trade-off. The bagged-boosting scheme is used to improve this by collaborating the multi-model constituent learners at the different levels. The stability guide classification scheme is delivered after or during the process of boosting.

In the Gentle Boost ensembles, the Subbaging and Gentle Boost are compared regarding their ability in the real-world datasets which give the 40**%** generalization error decrease but the true ability to capture the data revealed through texture analysis for the protein detection of gel electrophoresis images which gives the 0.9773 AUROS improved performance as compared to AUOC of 0.9574 which is based on SVM recursive feature.Ensemble methods such as bagging, boosting is used to improve the performance of the learning algorithm. Bagging is used for some base procedures and boosting is used to reduce the bias of the base procedure model. When bagging and boosting are combined and designed in-training or prediction ensembles then they improve the speed and accuracy.

W. Hong et al. [9**]** propose a tree model which is based on CART analysis used by clinicians for stratification in AP and to identify the patients that benefit from the close surveillance. There is a drawback of the scoring systems that they restrict the values of clinical. In this paper, a decision model is developed for prediction of severe

acute pancreatitis (SAP) based on the analysis of Classification and Regression Tree (CART). It is developed by enrolling the 420 patients who have acute pancreatitis on which logistic regression and univariate analysis are used to determine the predictors which are associated with the severe acute pancreatitis. Then CART analysis is used to develop the simple tree model for the prediction of severe acute pancreatitis (SAP).

In order toaccess the model performance, receiver operating characteristic (ROC) is used which is a training sample that is applied to the test sample. From the result, it is concluded that by using the logistic regression analysis, main four predictors variables of the SAP are determined i.e. blood urea nitrogen (BUN), serum calcium, pleural effusion, and systematic inflammatory response syndrome (SIRS). A Tree Model which is developed by the analysis of CART determines the high and low risk of developing the SAP among the cohorts which conclude that high risk is 79.03% and low risk is 7.80%. It is also concluded from the paper that the area which comes under the ROC curve of the tree model was greater than the APACHE II score. The predicted accuracy of the tree model in the test sample under the area of ROC curve is 0.86.

F. Gunes, et al. [25]illustrate about the stacked ensemble models which are used to improve the prediction accuracy by enabling the average out noise from the diverse models and by enhancing the generalizable signal. Basically, the stacked ensemble methods combine the predictions from the various machine learning algorithms which use these predictions as inputs for the second-level learning models. They described the method to generate the diverse set of the models by different techniques like gradient boosted decision trees, forest, logistic regression and factorization machines. After that, they will combine with the stacked-ensemble methods like gradient-boosting, hill climbing, and non-negative least squares in Machine learning and SAS Visual Data Mining.

The real-world applications of these methods create the big data problem which describes the using of stacked ensembles so that they produce the high prediction accuracy and the robustness. The used method is very powerful as it alters the initial mindset of data mining from the single best model in order to find the good complementary models. Due to the training of the many numbers of models and proper use of the cross validation increases the cost so as to avoid the overfitting.

They also describe efficiently handling of the computational expense in a modem SAS environment and to manage the ensemble workflow with the help of parallel computation in the distributed framework**.**

Table 2: Summarization of various evolutionary approaches

Authors	**Year**	**Objective**	**Method**	**Result**
A**.** S**.** Mohmed Salih, A**.** Abraham	2014	Evaluated the Ensembles design and combining different algorithms	ROC curves, Error Metrics, Specialty, Confusion Matrix, Cost/Benefit Methods	Novel Intelligent Ensemble method classifier achieved better outcomes
N**.** Arsov, M**.** Pavlovski, L**.** Basnarkov & L**.**	2017	Generating highly accurate prediction	Bagging and Boosting	Improved the speed and accuracy of ensembles**.**

Chapter 3

Problem Formulation

Irritable bowel syndrome abbreviated as IBS is one of the common disorder that impacts the larger intestine of a human being [1]. IBS caused approximately 53,000 deaths worldwide in 2013. The increasing prevalence of the disease is even heart-breaking. The symptoms of IBS are non-specific.

Along with this, its diagnosis is confirmed through invasive colonoscopy which is performed with consequent delays. Delayed pediatric IBS diagnosis leads to decreased growth and poor treatment results [2].

Wingfield, et al. (2016) investigated the feature relevance of feature set using the Boruta algorithm. The research uses two machine learning model including multilayer perceptrons (MLP) and Support vector machines (SVM) to determine the inflammatory bowel disease (IBD) presence, IBD activity from a stool sample and IBD subtype. The proposed method has various limitation including a focus on different activities and low accuracy.

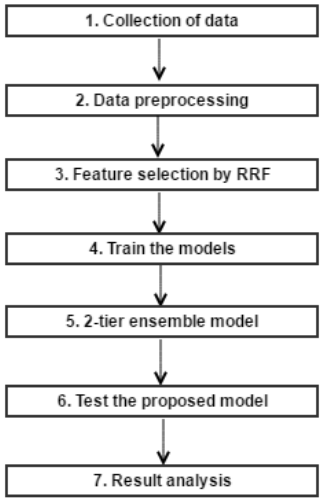


Figure 11: Flow chart to resolve the problem

The above flow diagram represents the process used to resolve the problem. We used ten different machine learning models for determining the properties that define the presence of IBS in the pediatric patient. The proposed method is accurate enough to detect the presence of IBS in the patient.

Chapter 4

Proposed Model for problem solution

This research is based on detection of Irritable Bowel Syndrome (IBS) using machine learning ensemble approach. The experimental analysis was done using ten different machine learning models: SVM, Neural Network, Linear Regression, Random Forest, Decision tree, AdaBoost (Adaptive Boosting), Naïve Bayes, Multilayer perceptron, Boosted tree, and Binary discrimination analysis.The data was collected from the website of UMASS Medical School. Here is the link to the website:http://escholarship.umassmed.edu/pediatrics_data/4/. The data was collectedfrom the pediatric patients to predict the presence of IBS in them.

4.1. Methodology

4.1.1. Proposed Model

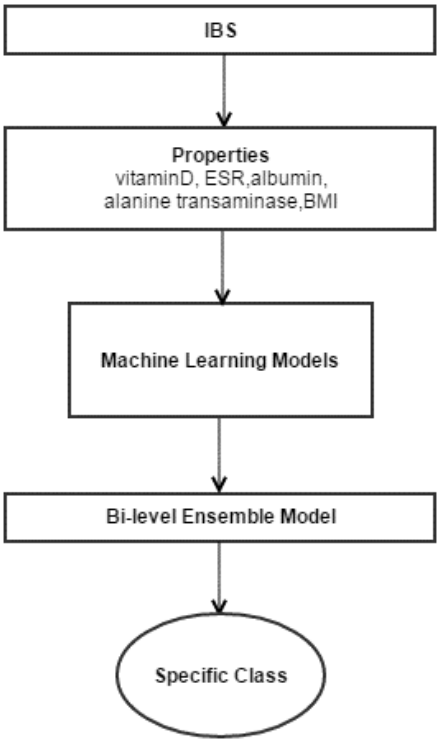


Figure 12: Flow chart of proposed model

The above figure demonstrates the flow chart of proposed model. The data of pediatric patients was collected to find the presence of IBS. The proposed model

analyzed the IBS data to discover the status of Vitamin D, ESR, Albumin, alanine transaminase, and Body Mass Index (BMI). Five different machine learning models were combined to propose an ensemble model to identify the specific class of IBS. We ensemble five different models to proposed our new model with efficient accuracy to predict whether a pediatric patient is IBS or not. The implementation of proposed ensemble model was done in R language. The RRF model was used for feature selection task.

4.1.2. Properties of IBS measured in this research

Irritable bowel syndrome is associated with the morbidity in the adolescents, and the children and the treatment options from the IBS are very limited. The persons who are suffering from the Irritable Bowel Syndrome (IBS), it is very uncomfortable and painful experience. It is categorized by the symptoms such as Constipation, Bloating, Abdominal Pain, Diarrhea, and discomfort.

Vitamin D

The potential solution or treatment of the Irritable bowel syndrome is by using the Vitamin D3. The Irritable bowel syndrome and Vitamin D3 gives the positive results for the patients who are suffering from Irritable bowel syndrome, and it is the effective and a natural treatment method for the treatment of the Irritable bowel syndrome patients. The supplementary high dose of the Vitamin D improved the IBS patients significantly. Vitamin D3 is the natural and effective way to treat the sufferers of Irritable bowel syndrome and improves the health of the patient [59].

The people who are suffering from the deficiency of the Vitamin D, it is essential to take the Vitamin D3 on aregular basis so that they can maintain the healthy body. And the people who are suffering from Irritable bowel syndrome, this is the more critical to the supplement of Vitamin D. It helps in supporting the immune function which is caused by the Irritable bowel syndrome and helps in the process of healing in the body of the person which is suffering from Irritable bowel syndrome [59].

Vitamin D has the beneficial effects of insomnia, anxiety, and low mood which are often seen in the patients which are suffering from Irritable bowel syndrome. The patients with IBS have an intolerance to certain types of foods which include the fatty substance. Vitamin D is the fat-soluble vitamin which is present in the high concentration in fatty fish which trigger the bowel spasms. So, by avoiding the consumptions can lead to the deficiency of the Vitamin D. the primary area for the absorption of the Vitamin D is in the gut and the erratic nature of Irritable bowel syndrome prevents vitamin D to be retained when the food is ingested. From the research, it is concluded that the supplements of the Vitamin D such as Vitamin D3 help in the multitude of the symptoms of Irritable bowel syndrome. The Vitamin D has contained the receptors in gut and brain which produce the brain chemical and serotonin which are experienced by the people who are suffering from Irritable bowel syndrome [59].

Erythrocyte sedimentation rate:

Erythrocyte sedimentation rate is also referred as the sed rate which is used to measure the amount of inflammation in the body. It is also used to monitor and detect cancer or inflammatory diseases. There is not a specific tool for the diagnostic by itself. For the diagnosis of Irritable bowel syndrome, Erythrocyte sedimentation rate is used which remove all the possible inflammatory causes of the abnormal digestive symptoms [60]. The normal range of the Erythrocyte sedimentation rate in the adults are given below:

- Men over 50 years old has less than 20 mm/hr ESR
- Men under 50 years old has less than 15 mm/hr ESR
- Women over 50 years old has less than 30 mm/hr ESR
- Women under 50 years old hasless than 20 mm/hr ESR

A normal patient of the Irritable bowel syndrome has the normal Erythrocyte sedimentation rate. The Erythrocyte sedimentation rate calculates the speed at which the mature red blood cells settled and used to screen the inflammatory disease. There is no need to test the Erythrocyte sedimentation rate if your blood tests and

thetemperature is normal, and your age is under 50, and the symptoms seen in you are similar to the Irritable bowel syndrome [60].

Erythrocyte sedimentation rate is aninexpensive, simple, and non-specific test which helps to detect the inflammation associated with the conditions such as autoimmune, cancer, and infectious diseases. It is the non-specific test due to its elevated results which indicate the presence of the inflammation. It can also be affected by the other conditions. it is used in the conjunctions with the other tests like C-reactive protein. It is also used to diagnose the inflammatory diseases, polymyalgia rheumatic, temporal arteritis, and systemic vasculitis. It is one of the tests which results in the diagnosis. It is also used to monitor the activities of the disease and then respond to therapy such as systemic lupus erythematosus [62].

Body Mass Index

Body mass index is a formula, which is used to relate the height of the body with weight. Basically, it is used by the health professionals and by the scientists to evaluate the implications of thehealth of being the certain weight for your height. The formula used to calculate the body mass index is given below:

Body mass index (BMI) = weight ÷ height

Body mass index is the number which is generated by dividing the weight which is in kilogram to the height which measures in meter squared. There is also an easy method to determine the body mass index by height in meters multiply by height in meters and after that divide the weight in kilograms by the calculated result [60].

Body mass index in the patients which are suffering from Irritable bowel syndrome related to the severity of symptom-related anxiety, abdominal pain, and bowel habit predominance. GI transit for the symptom pattern in the Irritable bowel syndrome patients. High BMI is related to the fast bowel transit which influences the symptoms in the patient with Irritable bowel syndrome. Body mass index in the Irritable bowel syndrome patients is associated differently which is based on the Irritable bowel syndrome bowel habit subtype. The Irritable bowel syndrome mix

subtype had the greater body mass index after controlling the all demographic information [62].

4.1.3. R programming

R is recognized as a programming language which is used by the programmers while developing the programs. It is a programming language which is used for the numerical analysis. It also used for the visuals representations and used to make the reporting in the programming functions. The R programming was created by the Ross Ihaka and the Robert Gentleman in the New Zealand. The R programming is currently developed by the R development fundamental team. The core of the R programming is made for the computer language which allows the dividing and the looping while developing the programming in the software of the computer.

The R programming uses the modular programming by performing the various functions. R programming also allows the additions in the programming with many procedures which is written in the C, C++, Python or the FORTRAN languages. It is used for the better efficiency of the programming and allows the functions in the better and the developed way.

The R language is freely available in the Public license. The R programming has also consisted of the binary versions, and it also provides the various operating systems like the Linux, Windows and the Mac. R programming is the free software which is present under the GNU- style,and it is also the part of the GNU project which is called the GNU S.

Evolution of the R Programming

The R language is introduced by the Ross Ihaka and the Robert Gentleman at the University of the Auckland at the branch of the statistics. The R language is introduced in the computer language for the first time in 1993.

- The R language is used by the individuals in the programming by sending the codes, and it also bugs the reports into the encoding system.

- In 1997 there was a core group introduced in the field of information technology which modifies the R language into the code documentation.

Features of the R programming

We studied earlier that the R programming language is used in programming the software of the computer for the graphical representations and the reporting. The following are some key features of the R programming which are helpful for developing the program in the various languages.

- R is the well-developed, modest and the operative programming language which uses looping and the branches. The R programming also includes the loops which perform the functions under the various programming procedures. It also provides the input and the output facilities so that the programming functions are performed effectively in each loop.
- R programming also has the effective data which handles all the programs in the encoding form, and it also facilitates the storage systems under the programming procedures.
- R programming also provides the group of the operators which helps to enable the various calculations in the programs. It also helps to allow the calculations in the arrays, lists, routes and the mediums.
- R programming also allows large and the combined collection of the tools for the data examination.
- R programming also provides the graphical facilities which help to analyze the data, and it also helps to display the data directly at the computer or the published papers.

So, the R programming allows the various features of the programming language and performs all the functions in the better and the effective way.

4.1.4. RF model

The RF model is the set of the model which is used in the software to evaluates the functions. It helps to represent the component in the programming. The

RF models measured the data at the numeric basis and computed the information to the time sphere response.

Each type of model is used for the different mathematical processes which help to represent the components in the numeric form.

RF model provides the prominent level of the component which helps to represent the data for the further usage after performing the detailed analyses with the use of RF circuit objects. The RF model mainly objects to the following things –

- It helps to compute the time domain figures which are essential for the RF components.
- It also exports the Verilog Models which performs the various functions in the RF components.

The RF models play the key role in the top to down design flows. The Top refers to the system which helps to specify the performance under the computer system while enabling the programming in the software. The structural specification describes the functions what the system should do in the programming not define the procedures to build the programmed in the operational procedures. In the RF model, the specification is the highest level of the concept. The RF models are specified at the top level and check the systems. On the other hand, the bottom of the designed flow in the RF model is helping to design the detailed diagram which enables the various functions in the one loop. The detailed design set in the RF model describes the functions to build the system in the RF process. It is also considered as a set of the diagrams in the device library which enables the layout to the operational systems in the RF model.

The detailed design in the RF model is abstracted at the lowest level and operates the system to evaluate into the numeric form. The bottom-up designs refer to the process which helps to design the system with the existing components of the RF models. The behavioural models in the RF system helpto design the space which is nearer to the top of the design flow. These parameters under the model affect the decisions which are handled at the top of the model, and it also ignored the other parameters. It is the first step in the top-down design of the RF systems which helps to choose the architectures to perform the functions efficiently.

The architecture in the interconnected diagrams of the RF model shows the sequences so that the operations are concerned in a systematic way. It also helps to transform the inputs and the outputs of the RF models.

The features of the RF blocks are specified by the RF model metrics such as the gain, sound figure and the third order interrupt points.

4.1.5. Machine learning technique with the R programming

In the scientific business and the business industries, there are a large amount of the data should be used. So, there is ahuge need foranalyses the data to make the correct decisions. But it is possible using the R programming which allows the users to be visualized the data and helps to run the data in the numerical form. But it is only possible by the applying the machine learning algorithm. So, the R programming is the best. The following are the reasons to enable the R programming in the machine learning algorithms-

- R programming provides the cutting-edge technology in the machine learning. The top researchers develop the machine learning by enabling the statistical methods in the R programming.
- It also helps to enable the new algorithms in the R programming, and it also added to the list of the packages.
- The R programming helps to install the packages in the machine learning processes and loads the packages to enables the data into the numeric form. It mainly includes the graphics and the statistical values.
- It also helps to run the codes in the command menu, but it does not easily allow to save the file and ensures to repeat and share the code in the machine learning.
- The functions of the R programming are generally introduced into the well-documented form. It contains both the details of whether it is related to the input or whether it is related to the output processes.
- It also ensures the accurate information in the machine learning algorithms to perform the functions accurately.

4.2. Models Specifications

The table below includes specifications of the Machine Learning modelsused in our ensemble model with their method names, required packages, and tuning parameters.

Table 3*:* Models Specifications

Model	Method	Required Package	Tuning Parameter
AdaptiveBoosting	ada	ada	None
Support vector machine (SVM)	ksvm	kernlab	kernel="rbfdot", type="C-svc"
Neural Network	nnet	nnet	size=10
Linear	Multinom	car	trace=FALSE, maxit=1000
Random Forest (RF)	rf	random forest	mtry=2, ntree=500
Decision Tree	rpart	None	usesurrogate=0, maxsurrogate=0

4.3. Parameters used to evaluate the performance of each machine learning model

4.3.1. Sensitivity

Sensitivity is also known as the true positive rate, theprobability of detection or recall. It is used to measure the proportions of the positives which are identified correctly such as the percentage of the ill people who are correctly identified as having the conditions [58]. It is also defined as the correctly identifying the presence of the diseases related to health. Sensitivity quantifies the avoiding of the false negatives. Basically, it refers to the testability which correctly detects the patients who do have the conditions [57]. For example, a medical test is used to define the disease and the sensitivity of the test detect the proportion of people who test positive for the

disease from those who have that disease. The mathematicalformula to calculate the sensitivity is given below:

$$sensitivity = \frac{numberofturepostives}{numberofturepositives + numberoffalsenegatives}$$

$$= \frac{numberofturepostives}{totalnumberofsickindividualsinpopulation}$$

$$= probabilityofapositivetestgiventhatthepatienthasthedisease$$

For rolling out the disease, a negative result in a test with high sensitivity is useful, and a high sensitivity test is only reliable when its results come out to be negative. Hence, they misdiagnose the disease from those who was suffering from it whereas the positive result having a high sensitivity is not useful for ruling in disease.

4.3.2. Specificity

Specificity is also known as the true negative rate which used to measure the proportion of negatives which correctly identify such as the percentage of the healthy people which are correctly identified as not having the conditions. It is also described as the probability of correctly identifying the presence of the diseases in the patients related to the heart. Specificity avoids the false positives [58]. Basically, specificity relates the ability of the test which correctly detects the patients without the conditions. For example, if there is a medical test for the diagnosing a disease then specificity of the test is proportional to the healthy patient who not having the disease and who will test negative for it [57]. The mathematical formula to calculate the specificity is given below:

$$specificity = \frac{numberofturenegatives}{numberofturenegatives + numberoffalsepositives}$$

$$= \frac{numberofturenegatives}{totalnumberofwellindividualsinpopulation}$$

$$= probabilityofanegativestestgiventhatthepatientiswell$$

For ruling in disease, then a positive result in a test having the high specificity is useful. This test gives the positive result in the patients who are healthy. The test which has 100% specificity then it will read negative and exclude the diseases accurately from the healthy patients. The positive result concludes the high probability of the presence of the disease.

4.3.3. Accuracy

The accuracy of the test is referredto the ability to differentiate the cases of healthy and unhealthy patients correctly. To determine the accuracy of the test, it is necessary to determine the proportions of true negatives and true positive in all the evaluated cases [58]. The accuracy of the test is calculated by using the following formula:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

4.3.4. Precision

Precision is defined as the close measurement which comes to another measurement. It is calculated by astatistical method which is known as the standard deviation. Standard deviation is related to how much, measurements, on average, which differ from each other. High standard deviation represents low precision whereas low standard deviation represents the high precision. It is also described as the random errors which measure the statistical variability [59]. The mathematical formula to find the predictive value is given below:

$$Precision = \frac{TP}{TP + FP}$$

4.3.5. F-Score

F-score is the statistical method which is used to calculate the accuracy of precision and recall. It also determines the accuracy which is based on the added and missed values. The F-Score is used to determine the performance of the model, and it

is used as the single measure of the accuracy of the model during the testing. It is also used to evaluate the weighted average of precision and recall. The recall is used to measure the quantity whereas the precision is used to measure the quality [57]. The mathematical formula to calculate the F-Score is given below:

$$F_{score} = 2\left(\frac{Precision * Recall}{Precision + Recall}\right)$$

When the value of F-score is 1, then it signifies the best score in terms of the prediction model and accuracy of the classification. And when the value of the F-Score is 0 then it signifies the worst score in terms of accuracy and prediction model.

4.4. Proposed Ensemble Model

In this research, we developed a novel combination of different machine learning models to develop an ensemble machine learning approach. This proposed ensemble model was developed to determine the presence of IBS in the pediatric patient. We used five machine learning models to develop ensemble model for detecting whether a pediatric patient has IBS or not. On the basis of the accuracy obtained by each model, we developed five ensemble models by combining three models in each ensemble. The five ensemble models are named as:

- E1 (AdaBoost, SVM, NN)
- E2 (SVM, NN, Linear regression)
- E3 (NN, Linear Regression, Random Forest)
- E4 (Ada, NN, Random Forest)
- E5 (SVM, NN, Random Forest)

The figure below shows the schematic diagram of our final combination of all five ensemble models.

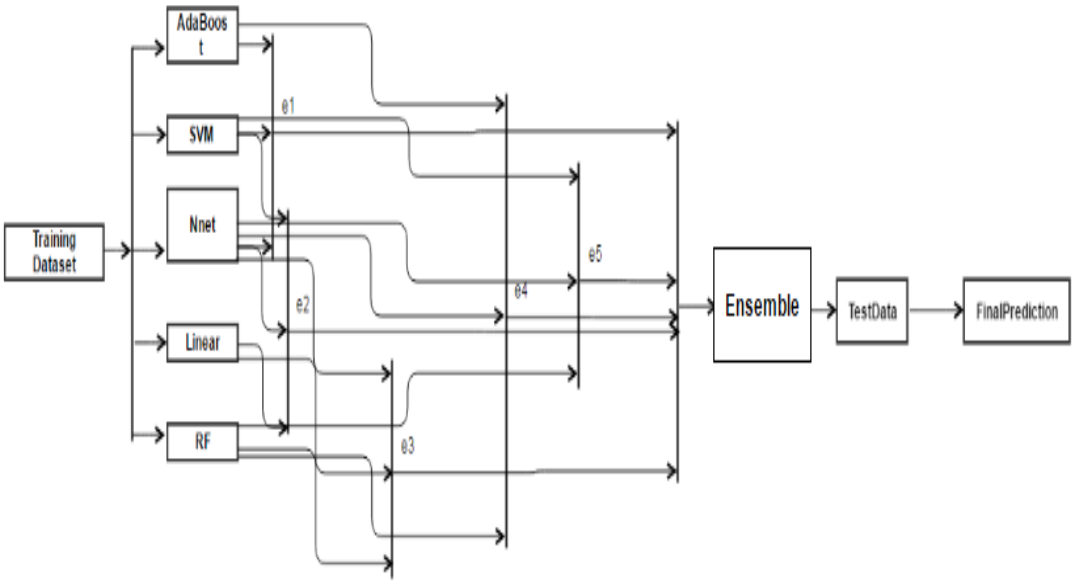


Figure 13: Final combination of ensemble models

Chapter 5

Experimental results

5.1. Plots of five ensemble models

This section includes the plots obtained by our experimental analysis of all ensemble models. The plotted graphs show the ROC, ROCH, H-measure, AUC, and Smoothed scored distribution values of each ensemble model.

Ensemble 1 (AdaBoost, SVM, NN)

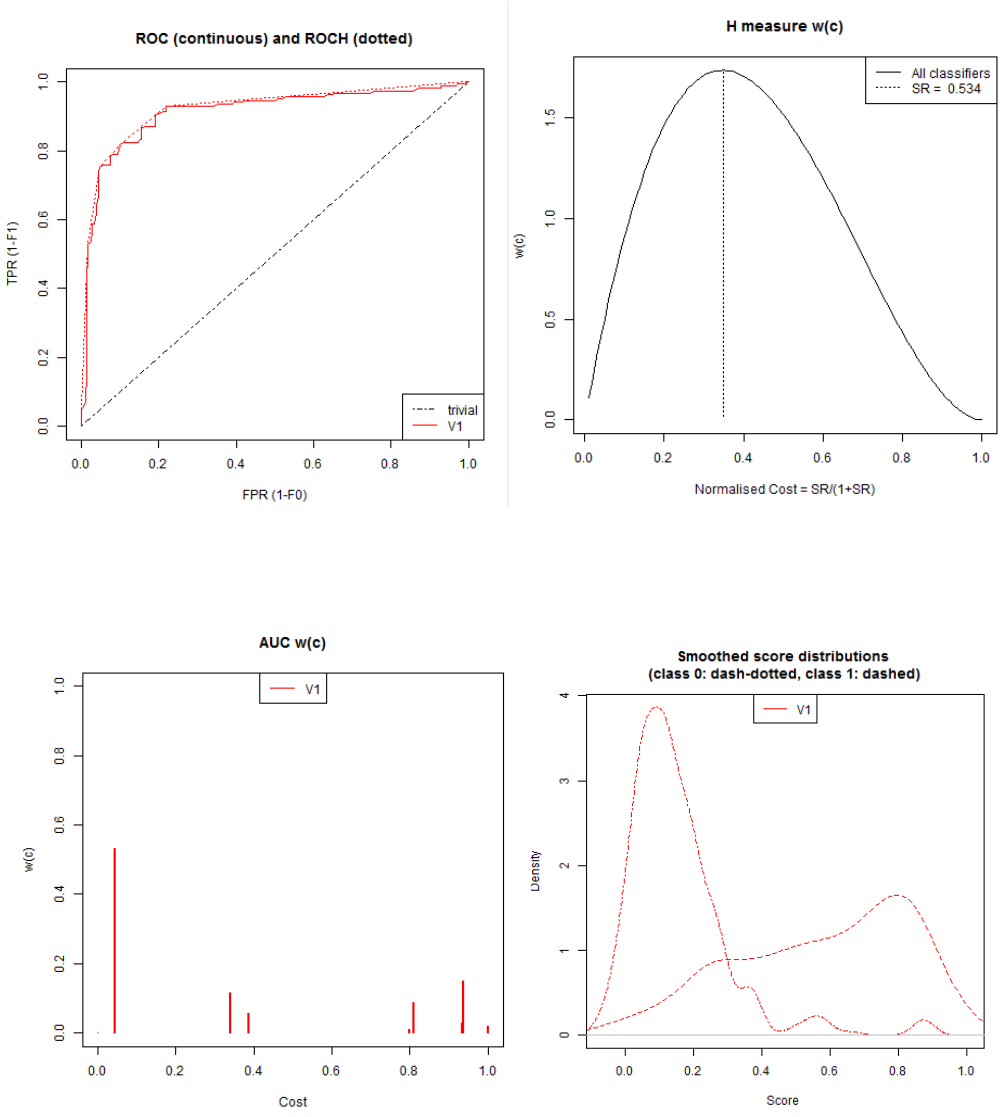


Figure 14*:* Results of Ensemble 1

Ensemble 2 (SVM, NN, Linear regression)

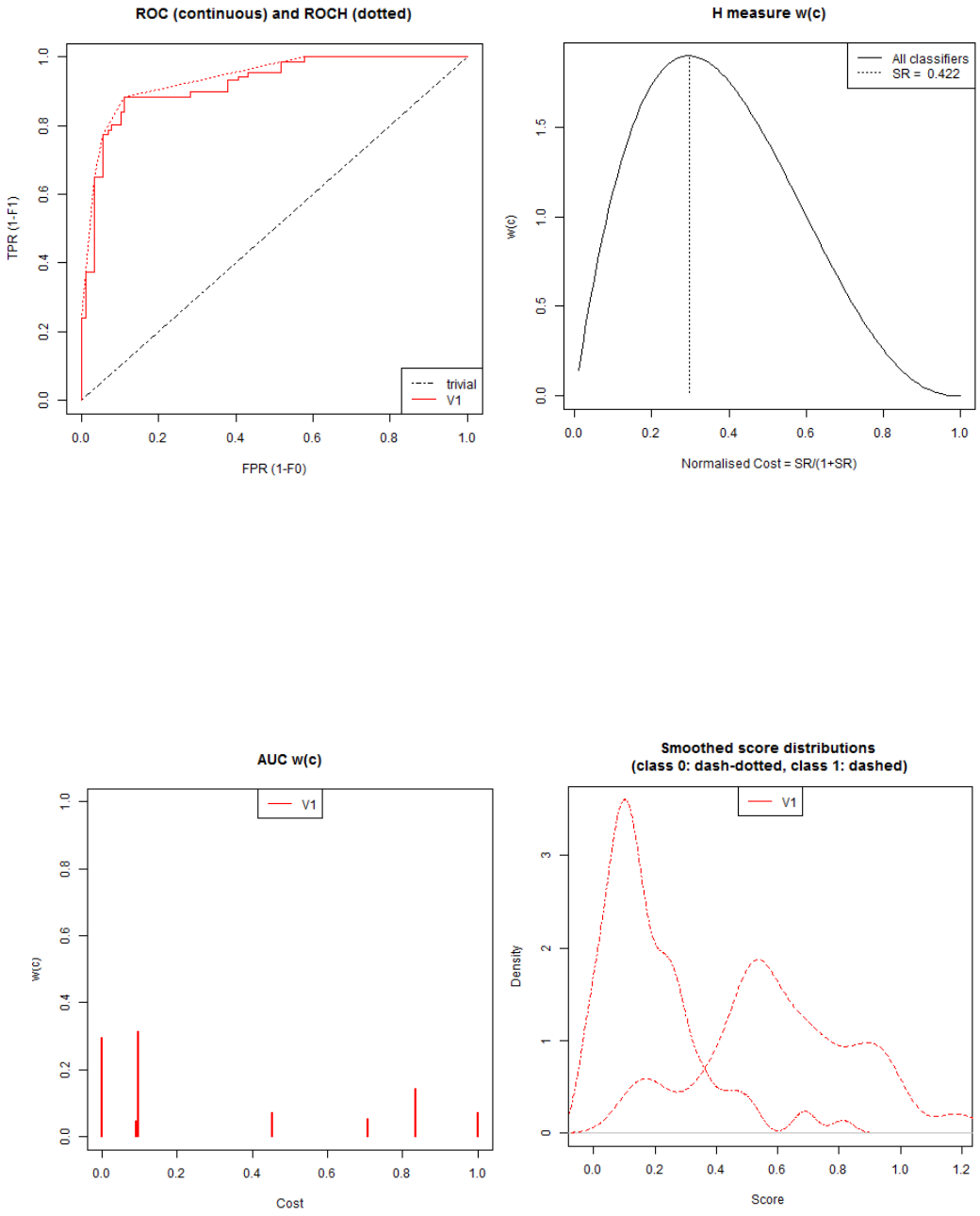
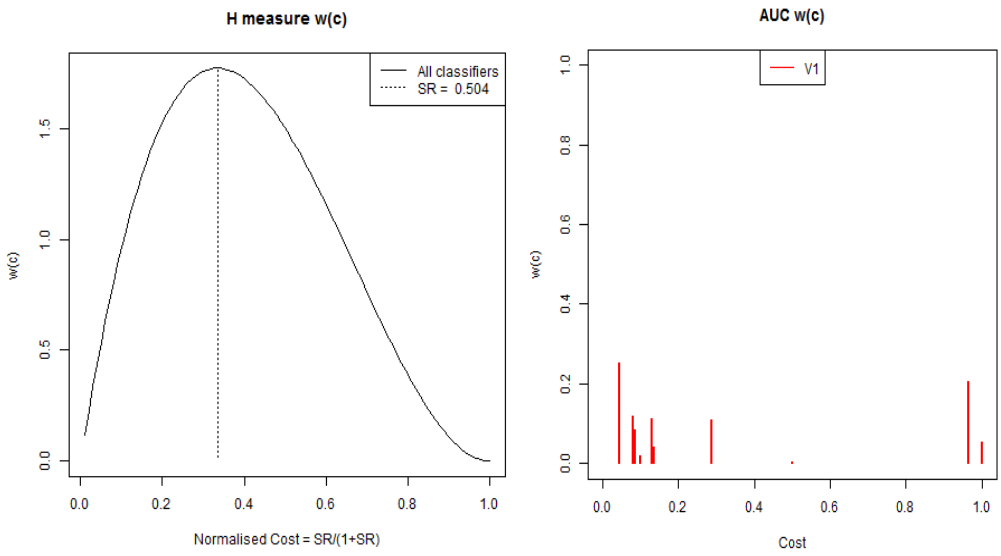


Figure 15: Results of Ensemble 2

Ensemble 3 (NN, Linear Regression, Random Forest)



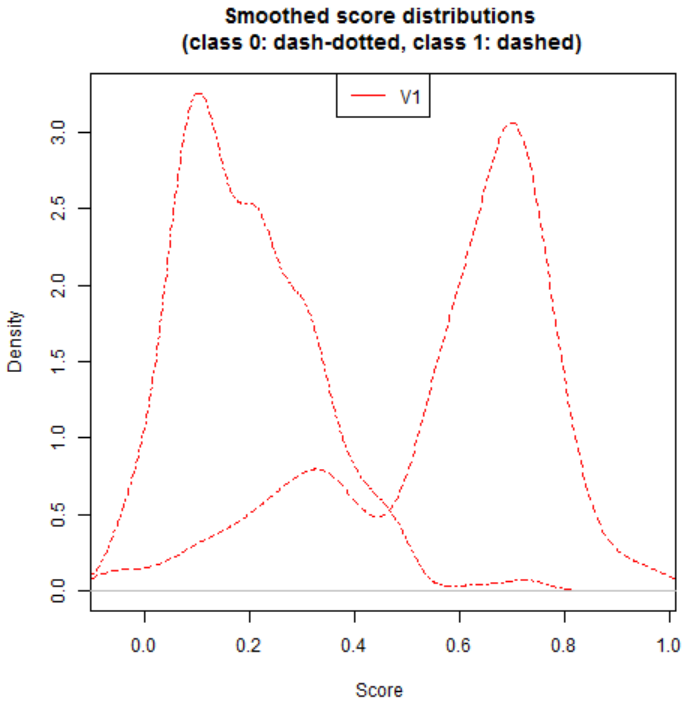


Figure 16*:*Results of Ensemble 3

Ensemble 4 (Ada, NN, Random Forest)

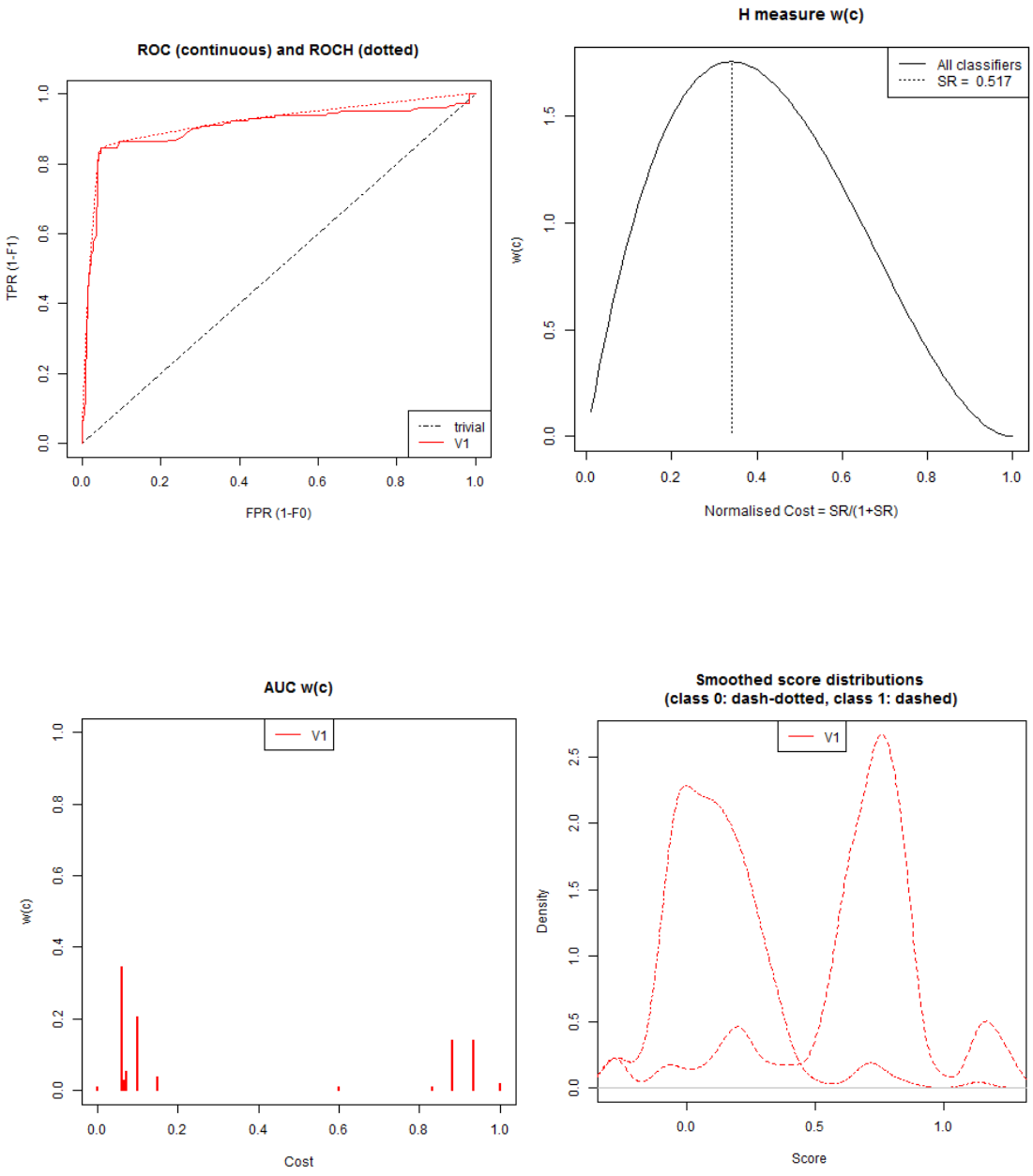


Figure 17*:* Results of Ensemble 4

Ensemble 5 (SVM, NN, Random Forest)

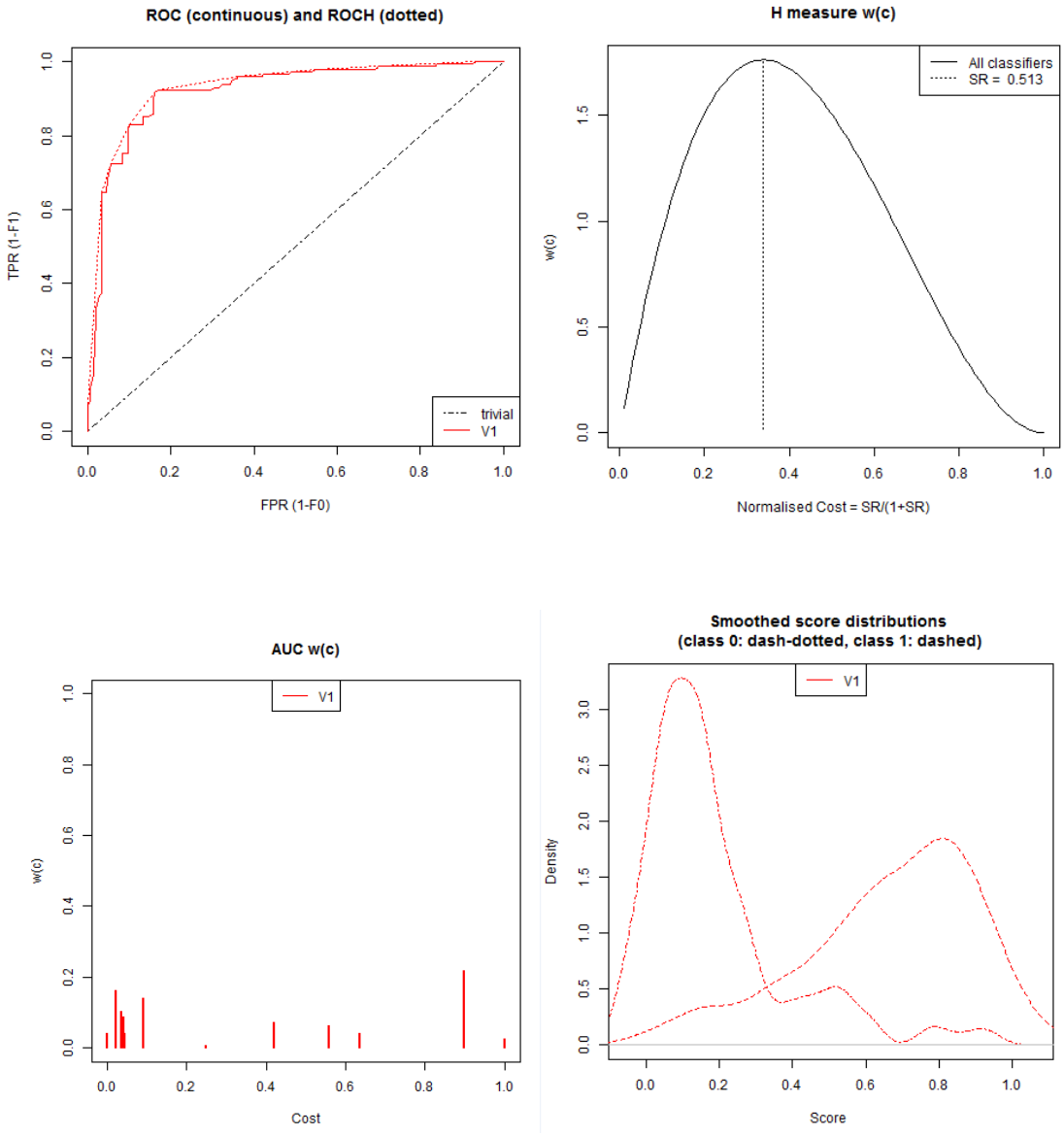


Figure 18: Results of Ensemble 5

The table below shows the values obtained for each model used in the experimental analysis.

Table 4*:* Values of each model

Models	Sensitivity	Specificity	H Measure w(c)	AUC	Accuracy
AdaBoost	0.678	0.909	0.478	0.832	82.17
SVM	0.555	0.927	0.446	0.825	79.92
Neural Network (NN)	0.562	0.898	0.354	0.784	78.21
Linear regression	0.522	0.843	0.316	0.777	73.04
Random Forest	0.708	0.833	0.475	0.823	81.52
Decision Tree	0.459	0.862	0.159	0.659	68.72
Naïve Bayes	0.402	0.856	0.159	0.683	67.42
Multilayer perceptron	0.422	0.87	0.171	0.68	68.26
Boosted tree	0.451	0.853	0.236	0.745	71.51
Binary discrimination analysis	0.515	0.795	0.17	0.685	68.84

From the above table, we can see the overall accuracy obtained by each machine learning model. AdaBoost and Random Forest models achieved thehighest

accuracy of 82.17 % and 81.52%, respectively, as compared to other eight models. Apart from these two models, the accuracy achieved by SVM is 79.92%, Neural Network as 78.21%, and Linear regression as 73.07%. The Naïve Bayes model achieved the least accuracy of 68.26% as compared to other models.

The table below shows the values obtained for five groups made by combining three models.

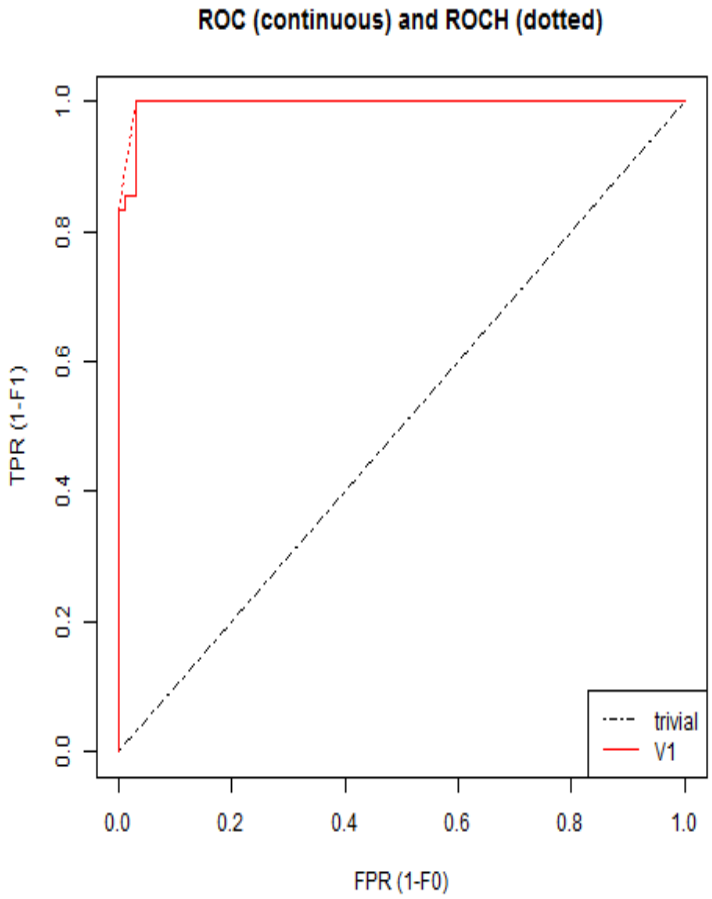
Table 5*:* Values of five ensemble models

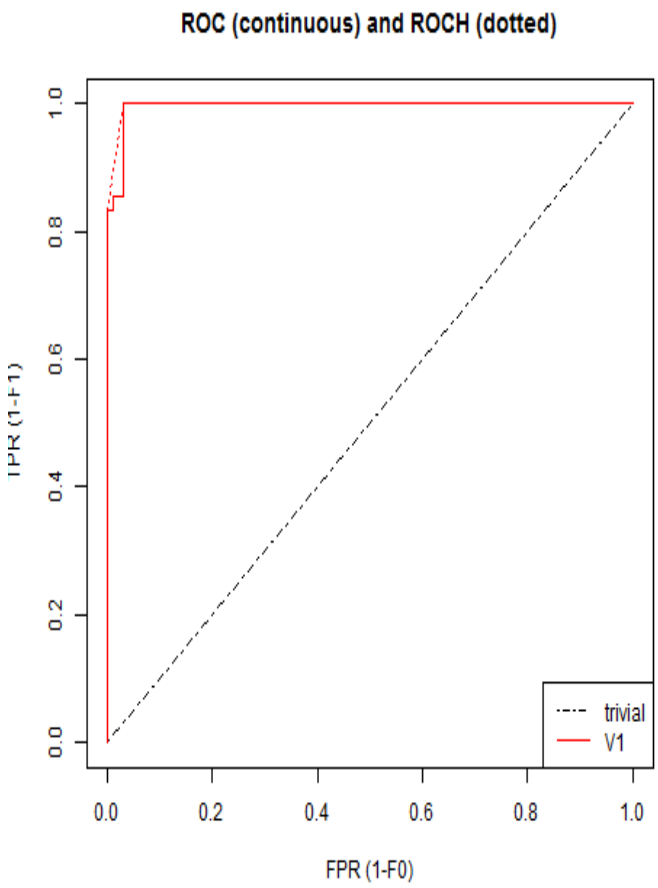
Models Group	Sensitivity	Specificity	H Measure w(c)	AUC	Accuracy
E1 (AdaBoost, SVM, NN)	0.615	0.96	0.651	0.912	86.41
E2 (SVM, NN, Linear regression)	0.701	0.945	0.672	0.921	83.55
E3 (NN, Linear Regression, Random Forest)	0.733	0.989	0.677	0.91	82.5
E4 (Ada, NN, Random Forest)	0.831	0.958	0.713	0.903	81.01
E5 (SVM, NN, Random Forest)	0.741	0.97	0.689	0.882	88.64

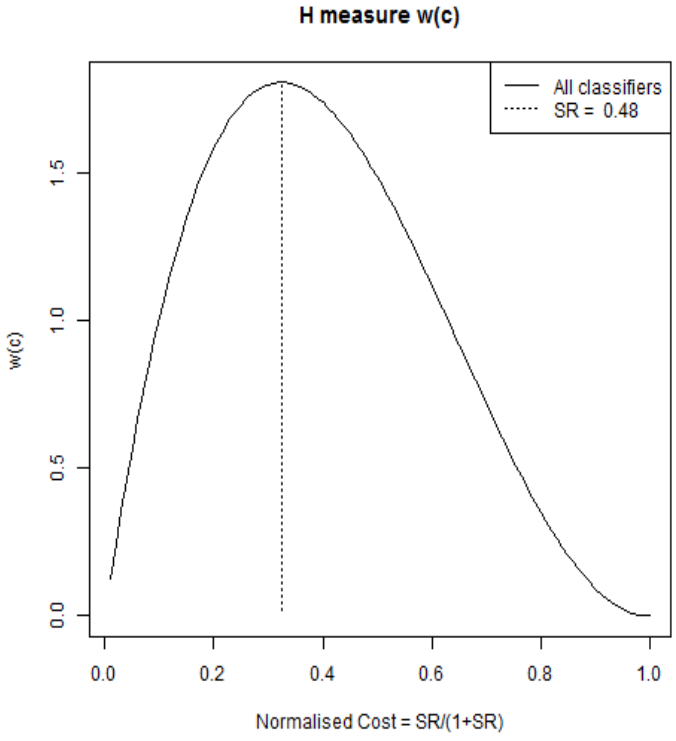
From the above table, we can see the overall accuracy obtained by each ensemble model. Ensemble 5 (E5) combining SVM, NN, and Random Forest, obtained the highest accuracy of 88.64% as compared to other models. Ensemble 1 (E1) combining AdaBoost, SVM, and Random Forest, obtained the accuracy of 86.41%. Whereas other ensemble models (E2, E3, and E4) achieved the accuracy of 83.55%, 82.5%, and 81.01%, respectively. By combining all five ensemble models with high accuracies, we created a final combination of ensemble models. Chapter 5 presented the results of our final combination.

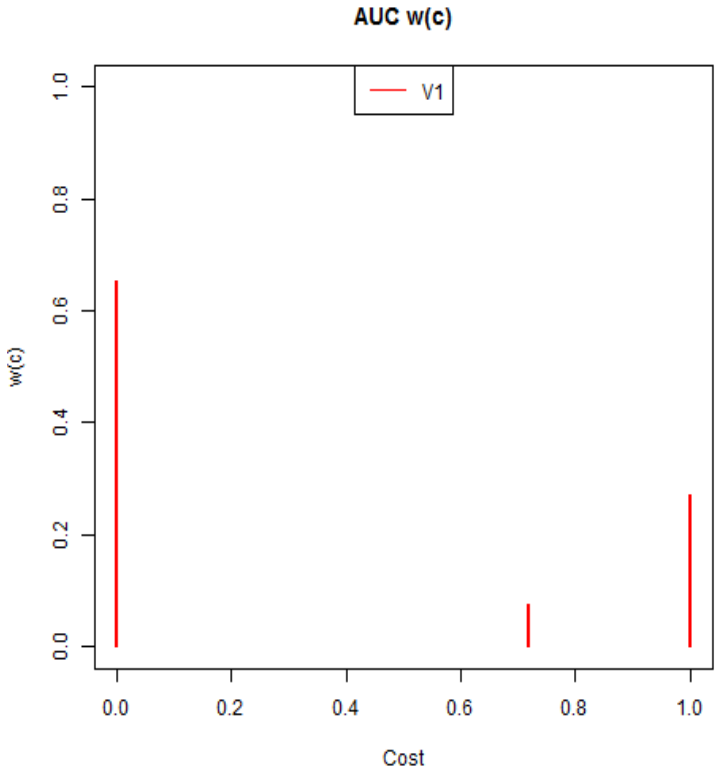
5.2. Final combination results

This section includes the plots obtained by our final combination of all ensemble models. The plotted graphs show the ROC, ROCH, H-measure, AUC, and Smoothed scored distribution values our final ensemble model.




ROC (continuous) and ROCH (dotted)
TPR (1-F1)
0.0
0.2
0.4
0.6
0.8
1.0
trivial
V1
FPR (1-F0)


H measure w(c)
w(c)
0.0
0.5
1.0
1.5
All classifiers
SR = 0.48
0.2
0.4
0.6
0.8
Normalised Cost = SR/(1+SR)



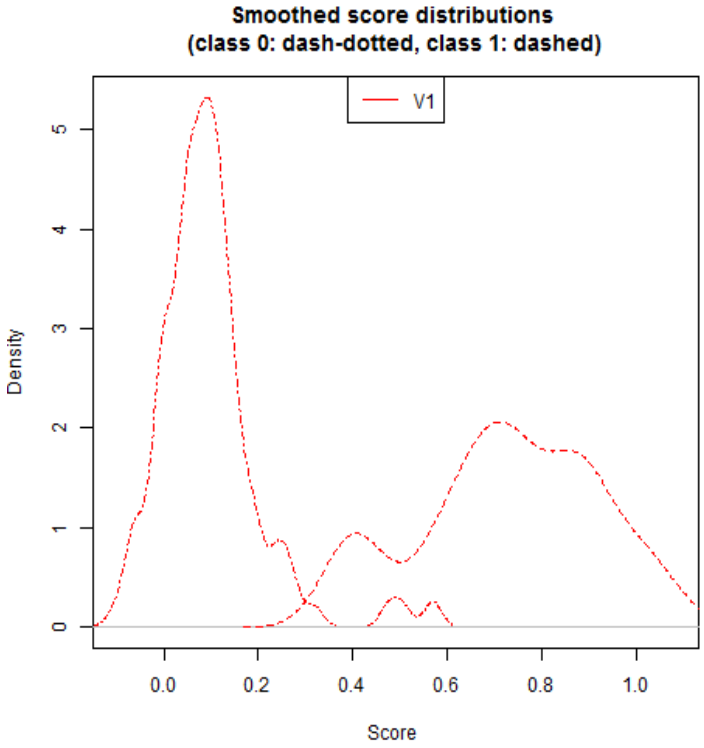


Figure 19: Results of Final combination

The graph below shows the plotting of accuracy of our final combination ensemble model at different runs. The result of our research shows that the proposed method is 93.32% accurate to detect the presence of IBS.

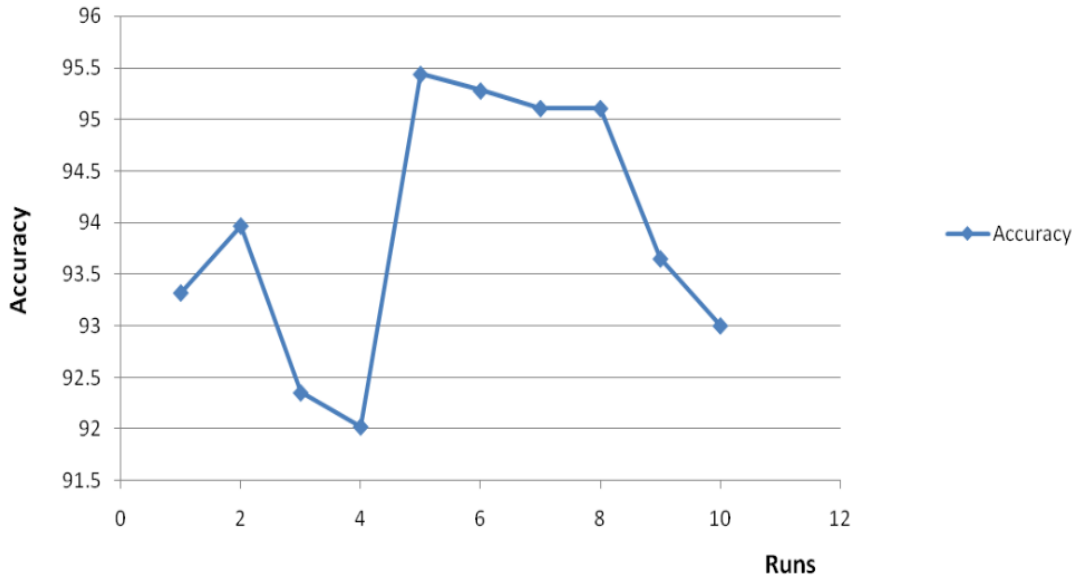


Figure 20*:* Accuracy vs. Runs of final ensemble model

Chapter 6

Conclusion and Future Scope

IBS is the common disorder that impacts the larger intestine of a human being. It does not cause any abnormality in the structure, but likely to disrupt the function of a part of the body. In simple words, even if we see the gut under the microscope, we will not see any structural problem in the gut. All parts would look working normally. But it would cease to function appropriately. There is no proper way to detect and diagnose the IBS. Unfortunately, the prevalence of the disease is increasing day by day.

In this research, we presented a novel combination of different machine learning models to develop anensemble machine learningapproach that can effectively determine the presence of IBS in thepediatric patient. We used different machine learning models to develop an ensemble approach for detecting whether a pediatric patient has IBS or not. The result of our research shows that the proposed method is 93.32% accurate to detect the presence of IBS.

As limited research has been done in predicting the IBS. The proposed method has the ability to assist other researchers in the same area. Future work will include using more number of features to increase the accuracy in predicting the IBS.

References

[1]. Nwosu, B. U., Maranda, L., & Candela, N. "Vitamin D status in pediatric irritable bowel syndrome".PLOS ONE, vol. 12, No. 2, 2017.

[2]. Wingfield, B., Coleman, S., McGinnity, T., & Bjourson, A. J."A metagenomic hybrid classifier for pediatric inflammatory bowel disease."2016 International Joint Conference on Neural Networks (IJCNN), pp. 1-7, 2016.

[3]. Salih, A. S., & Abraham, A. "Novel Ensemble Decision Support and Health Care Monitoring System."Journal of Network and Innovative Computing, vol.2, pp. 41-51, 2016.

[4]. Dietterich, T. G. "Ensemble Methods in Machine Learning."Multiple Classifier Systems, pp. 1-15,2000.

[5]. Cosgun, E., Limdi, N. A., & Duarte, C. W. "High-dimensional pharmacogenetic prediction of a continuous trait using machine learning techniques with application to warfarin dose prediction in African Americans."Bioinformatics, vol. 27, No. 10, 1384-1389, 2011.

[6]. Penny, K. I., & Smith, G. D. "The use of data-mining to identify indicators of health-related quality of life in patients with irritable bowel syndrome."Proceedings of the ITI 2009 31st International Conference on Information Technology Interfaces, vol.21, pp. 2761-2771, 2009.

[7]. Nidhi, N., Glick, M., Davies, J. W., & Jenkins, J. L. "Prediction of Biological Targets for Compounds Using Multiple-Category Bayesian Models Trained on Chemogenomics Databases."ChemInform, vol. 37, No. 30, 1124-1133, 2006.

[8]. Arsov, N., Pavlovski, M., Basnarkov, L., & Kocarev, L. "Generating highly accurate prediction hypotheses through collaborative ensemble learning."Scientific Reports, vol. 7, pp. 44-49, 2017.

[9]. Hong, W., Dong, L., Huang, Q., Wu, W., Wu, J., & Wang, Y. "Prediction of Severe Acute Pancreatitis Using Classification and Regression

Tree Analysis.”Digestive Diseases and Sciences, vol. 56, No. 12, pp. 3664-3671, 2011.

[10]. Kurt, I., Ture, M., &Kurum, A. T. “Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease.”Expert Systems with Applications, vol. 34, No. 1, pp. 366-374, 2008.

[11]. RAZI, M., & ATHAPPILLY, K. “A Comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models.”Expert Systems with Applications, vol. 29 No.1, pp. 65-74, 2005.

[12]. Awad, R. A., Martin, J., Guevara, M., Ramos, R., Noguera, J. L., Camacho, S., … Toriz, A. “Defaecography in patients with irritable bowel syndrome and healthy volunteers.”International Journal of Colorectal Disease, vol.12, No. 2, pp. 91-94, 1997.

[13]. "Irritable Bowel Syndrome: How to Deal with It Naturally,"Beyonddiet.com, 2017. [Online]. Available: https://www.beyonddiet.com/articles/342/irritable-bowel-syndrome-how-to-deal-with-it-natur. [Accessed: 27- May- 2017].

[14]. "IBS Symptoms | IBS Treatment Center", Ibstreatmentcenter.com, 2017. [Online]. Available: https://ibstreatmentcenter.com/ibs-symptoms. [Accessed: 27- May- 2017].

[15]. "2.1. General structure — ANNarchy 4.6.2 documentation", Annarchy.readthedocs.io, 2017. [Online]. Available: http://annarchy.readthedocs.io/en/stable/manual/Structure.html. [Accessed: 27- May- 2017].

[16]. "Support Vector Machine without tears", Digg Data, 2017. [Online]. Available: http://diggdata.in/post/94066544971/support-vector-machine-without-tears. [Accessed: 27- May- 2017].

[17]. J. Zia, C. Chung, J. Schroeder, S. Munson, J. Kientz, J. Fogarty, E. Bales, J. Schenk and M. Heitkemper, "The feasibility, usability, and clinical

utility of traditional paper food and symptom journals for patients with irritable bowel syndrome,"Neurogastroenterology & Motility, vol. 29, no. 2, p. e12935, 2016.

[18]. R. Soares, "Irritable bowel syndrome: A clinicalreview,"World Journal of Gastroenterology, vol. 20, no. 34, p. 12144, 2014.

[19]. O. Grundmann, "Complementary and alternative medicines in irritable bowel syndrome: An integrative view,"World Journal of Gastroenterology, vol. 20, no. 2, pp. 346, 2014.

[20]. Q. Tang, "Cognitive-behavioral therapy for the management of irritable bowel syndrome,"World Journal of Gastroenterology, vol. 19, no. 46, p. 8605, 2013.

[21]. T. MAJID, "Quantitative analysis on the level of IBS acceptance in the Malaysian construction industry,"Journal of Engineering Science and Technology, vol. 6, no. 2, pp. 179 - 190, 2011.

[22]. R. Sood, G. Law and A. Ford, "Diagnosis of IBS: symptoms, symptom-based criteria, biomarkers or 'psychomarkers'?", Nature Reviews Gastroenterology & Hepatology, vol. 11, no. 11, pp. 683-691, 2014.

[23]. V. Shankar, N. Reo, and O. Paliy, "Simultaneous fecal microbial and metabolite profiling enables accurate classification of pediatric irritable bowel syndrome,"Microbiome, vol. 3, no. 1, 2015.

[24]. D. Saulnier, K. Riehle, T. Mistretta, M. Diaz, D. Mandal, S. Raza, E. Weidler, X. Qin, C. Coarfa, A. Milosavljevic, J. Petrosino, S. Highlander, R. Gibbs, S. Lynch, R. Shulman and J. Versalovic, "Gastrointestinal Microbiome Signatures of Pediatric Patients with Irritable Bowel Syndrome", Gastroenterology, vol.141, no. 5, pp. 1782-1791, 2011.

[25]. F. Güneş, R. Wolfinger and P. Tan, "Stacked Ensemble Models for Improved Prediction Accuracy,"SAS Institute Inc, pp. 1-19, 2017.

[26]. E. Giannetti and A. Staiano, "Probiotics for Irritable Bowel Syndrome: Clinical Data in Children,"JPGN, vol. 63, pp. S25-S26, 2016.

[27]. G. Longstreth, "Symptoms and Tests for Irritable Bowel Syndrome: Diagnosing a Complex Disorder,"Clinical Gastroenterology and Hepatology, vol. 8, no. 2, pp. 132-136, 2010.

[28]. S. Park, K. Han and C. Kang, "Relaxation Therapy for Irritable Bowel Syndrome: A Systematic Review,"Asian Nursing Research, vol. 8, no. 3, pp. 182-192, 2014.

[29]. T. Wilkins, C. PEPITONE, B. ALEX and R. Schade, "Diagnosis and Management of IBS in Adults,"American Family Physician, vol. 86, no. 5, pp. 419-426, 2012.

[30]. M. Mustafa, J. Menon, and R. Muniandy, "Irritable Bowel Syndrome: Pathophysiology, Management, and Treatment,"Journal of Dental and Medical Sciences, vol. 14, no. 6, pp. 70-76, 2015.

[31]. H. Vahedi, R. Ansari, M. Mir-Nasseri and E. Jafari, "Irritable Bowel Syndrome: A Review Article,"Middle East Journal of Digestive Diseases, vol. 2, no. 2, pp. 66-77, 2010.

[32]. M. McOmber and R. Shulman, "Recurrent abdominal pain and irritable bowel syndrome in children,"Current Opinion in Pediatrics, vol. 19, no. 5, pp. 581-585, 2007.

[33]. S. Magge and J. Wolf, "Complementary and alternative medicine and mind–body therapies for thetreatment of irritable bowel syndrome in women,"Women's Health, vol. 9, no. 6, pp. 557-567, 2013.

[34]. N. Ragavan, S. Kumar, T. Chye, S. Mahadeva and H. Shiaw-Hooi, "Blastocystis sp. in Irritable Bowel Syndrome (IBS) - Detection in Stool Aspirates during Colonoscopy", PLOS ONE, vol. 10, no. 9, p. e0121173, 2015.

[35]. F. Dogruman-Al, Z. Simsek, K. Boorom, E. Ekici, M. Sahin, C. Tuncer, S. Kustimur and A. Altinbas, "Comparison of Methods for Detection of Blastocystis Infection in Routinely Submitted Stool Samples, and also in IBS/IBD Patients in Ankara, Turkey,"PLoS ONE, vol. 5, no. 11, p. e15484, 2010.

[36]. S. Rajindrajith and N. M. Devanarayana, “Subtypes and Symptomatology of Irritable Bowel Syndrome in Children and Adolescents: A School-based Survey Using Rome III Criteria,” Journal of Neurogastroenterology and Motility, vol. 18, no. 3, pp. 298–304, 2012.

[37]. Y. Tanaka and M. Kanazawa, “Biopsychosocial Model of Irritable Bowel Syndrome,” Journal of neurogastroenterology and motility, vol. 17, no. 2, pp. 131–139, Apr. 2011.

[38]. A. Abraham, A. Salih, and M. Salih, "Novel Ensemble Decision Support and Health Care Monitoring System," Journal of Network and Innovative Computing, vol. 2, pp. 041–051, 2014.

[39]. M. Andoh, Y. Sato, H. Sakamoto, T. Yoshida, and M. Ohtaki, "Detection of inappropriate samples in association studies by an IBS-based method considering linkage disequilibrium between genetic markers," Journal of Human Genetics, vol. 55, no. 7, pp. 436–440,2010.

[40]. S. P. Paul, P. Barnard, C. Bigwood, and D. C. A. Candy, "Challenges in themanagement of irritable bowel syndrome in children," Indian Pediatrics, vol. 50, no. 12, pp. 1137–1143, 2013.

[41]. J. M. Hollier, D. I. Czyzewski, M. M. Self, E. M. Weidler, E. O. B. Smith, and R. J. Shulman, "Pediatric Irritable Bowel Syndrome Patient and Parental Characteristics Differ by Care Management Type," Journal of Pediatric Gastroenterology and Nutrition, vol. 64, no. 3, pp. 391–395, 2017.

[42]. A. Hegland, E. Winjum, P. Spilling, C. Rong, and O. Kure, "Analysis of IBS for MANET Security in Emergency and Rescue Operations," 20th International Conference on Advanced Information Networking and Applications - Volume 1 (AINA06), 2006.

[43]. R. Ikechi, B. Fischer, J. Desipio, and S. Phadtare, "Irritable Bowel Syndrome: Clinical Manifestations, Dietary Influences, and Management," Healthcare, vol. 5, no. 2, p. 21, 2017.

[44]. E. Chiou and S. Nurko, "Functional abdominal pain and irritable bowel syndrome in children and adolescents," Therapy, vol. 8, no. 3, pp. 315–331, 2011.

[45]. M. Fadgyas-Stanculete, A.-M. Buga, A. Popa-Wagner, and D. L. Dumitrascu, "The relationship between irritable bowel syndrome and psychiatric disorders: from molecular changes to clinical manifestations," Journal of Molecular Psychiatry, vol. 2, no. 1, p. 4, 2014.

[46]. A. M. Yeh, A. Wren, and B. Golianu, "Mind–Body Interventions for Pediatric Inflammatory Bowel Disease," Children, vol. 4, no. 4, p. 22, Mar. 2017.

[47]. S. Ballou and L. Keefer, "Psychological Interventions for Irritable Bowel Syndrome and Inflammatory Bowel Diseases," Clinical and Translational Gastroenterology, vol. 8, no. 1, 2017.

[48]. L. E. Miller, "Study design considerations for irritable bowel syndrome clinical trials," Annals of Gastroenterology, vol. 27, pp. 338–345, 2014.

[49]. Y. Tse, D. Armstrong, C. N. Andrews, A. Bitton, B. Bressler, J. Marshall, and L. W. C. Liu, "Treatment Algorithm for Chronic Idiopathic Constipation and Constipation-Predominant Irritable Bowel Syndrome Derived from a Canadian National Survey and Needs Assessment on Choices of Therapeutic Agents," Canadian Journal of Gastroenterology and Hepatology, vol.2017, pp. 1–11, 2017.

[50]. E. A. Mayer, J. S. Labus, K. Tillisch, S. W. Cole, and P. Baldi, "Towards a systems view of IBS," Nature Reviews Gastroenterology & Hepatology, vol. 12, no. 10, pp. 592–605, 2015.

[51]. Mennitt, D., Sherrill, K., & Fristrup, K. "A geospatial model of ambient sound pressure levels in the contiguous United States."The Journal of the Acoustical Society of America, 135(5), 2746-2764, 2014.

[52]. T. Kim, "Decision Forests and discriminant analysis," Slideshare.net, 2017. [Online]. Available: https://www.slideshare.net/potaters/decision-forests-and-discriminant-analysis. [Accessed: 03- Jun- 2017].

[53]. J. Grisanti, "Decision Trees: An Overview | Aunalytics", Aunalytics.com, 2017. [Online]. Available: http://www.aunalytics.com/decision-trees-an-overview/. [Accessed: 03- Jun- 2017].

[54]. Tutorialspoint. (2016). R programming. Retrieved from http://www.tutorialspoint.com/r/r_tutorial.pdf

[55]. "RF Analog Impairments Description and Modeling."RF Analog Impairments Modeling for Communication Systems Simulation, pp. 37-105, 2012.

[56]. Chang, A. (2016). R for Machine Learning. R for Machine Learning. Retrieved from https://ocw.mit.edu/courses/sloan-school-of-management/15-097-prediction-machine-learning-and-statistics-spring-2012/lecture-notes/MI

[57]. H., K., H., J., & J., G. "Diagnosing Coronary Heart Disease using Ensemble Machine Learning."International Journal of Advanced Computer Science and Applications, vol. 7, No. 10, pp. 30-39, 2016.

[58]. Adjorlolo, S. (2016). Diagnostic Accuracy, Sensitivity, and Specificity of Executive Function Tests in Moderate Traumatic Brain Injury in Ghana.Assessment.

[59]. Anderson, P. (2017). Accuracy and Precision Tutorial | Sophia Learning. Retrieved from https://www.sophia.org/tutorials/accuracy-and-precision--3

[60]. Glutathionepro. (2013). New Study: Vitamin D3 and IBS Treatment - Glutathione Pro. Retrieved from http://glutathionepro.com/new-study-vitamin-d3-and-ibs-treatment/

[61]. Lerche, O., & Thistlethwaite, F. (2016). The cure for IBS discovered, and it's just natural vitamin D | Health | Life & Style | Express.co.uk. Retrieved from http://www.express.co.uk/life-style/health/634625/IBS-cured-Vitamin-D-alzheimer-s-dementia-sunshine

[62]. Labtestonline. (2014). Erythrocyte Sedimentation Rate (ESR): The Test | Erythrocyte Sedimentation Rate, ESR Test: Erythrocyte Sedimentation Rate Test; Sed Rate Test | Lab Tests Online. Retrieved from https://labtestsonline.org/understanding/analytes/esr/tab/test/

List of Publications and video Link

[1] A. Jat, M. Kaur, " Prediction of Pediatric IBS using Machine Learning Models," *ISRD- International Conference on Current Resaecrh in Enginerring and Technology*, 2017 (Accepted).

[2] https://www.youtube.com/watch?v=bxyrujMgYG8(Video Link)