A Comparative Analysis using Machine Learning Approach for Sleep Apnea Disorder Mining

V. Ponmalar[1], B. Priyadharsini[2], S. Sakthivel[3], S. Savitha[4], K. Logeswaran[5]

[1,2,3]Student, Department of Computer Science and Engineering, K.S.R. College of Engineering, Tiruchengode-637215, Tamilnadu, India. Email:ponnu231198@gmail.com[1], priyabaskar10498@gmail.com[2],sakthivelhero1999@gmail.com[3].

[4]Assistant Professor, Department of Computer Science and Engineering, K.S.R. College of Engineering, Tiruchengode-637215, Tamilnadu, India. Email:infosavi@gmail.com.

[5]Assistant Professor, Department of Information Technology, Kongu Engineering College, Perundurai-638060, Tamilnadu, India. Email:klogesbtechit@gmail.com.

Abstract - Sleep Apnea is a discomfort disease caused by Stress. It is typically found in prevalent disorder peoples. For long periods of time, Authorities are trying to find out some of the features on Sleep Apnea illness so that they can rightly categorize sleep sickness because different sleep disorder requires different Cases of treatment. India has been targeted to Sleep Apnea disease from the last few years. Sleep Apnea is used in classification techniques to evaluate and compare their performance. Java with Weka was used as a Data mining tool for the classification of data. Firstly, we will assess the presentation of all the techniques separately with the help of tables and graphs depending upon the dataset and secondly, we will compare the presentation of all the techniques.

Keywords: Sleep Apnea, Apnea classification; Naïve Bayes; J48; SMO; REP.

I. INTRODUCTION

Sleep Apnea infection is vital to disease caused by Sleep Apnea germ, which extends in body of humans by Lungs airways [1]Sleep Apnea infection has threatened a billion populations all around the world [2]. India has been victim of this rapidly growing disease from the last few years. Since 2007 in India, large number of cases was marked especially in Secunderabad. In 1994 at UP India's first case of Sleep Apnea was appeared and Sleep Apnea's outbreak in 2011, which was more life-discomfort than previous years and 1400 people were affected [3]Sleep Apnea is divided into two Case s, i.e., Case 1 and Case 2, according to world health organization [3]First one is classical Sleep Apnea called Sleep Apnea and the other is Sleep Apnea hemorrhagic Apnea. SP1, SP2, SP3, and SP4 are further four Cases of Sleep Apnea hemorrhagic Apnea. SP is revealed by the start of Apnea which continues for 2 to 7 days with the number of signs like leakage of plasma, shock and weak pulse. In the earliest cases it's hard to differentiate Sleep Apnea from Sleep Apnea hemorrhagic Apnea. Different techniques for Sleep Apnea classification can be used such as NB classifier; decision tree, KNN Technique, multilayered Technique and SVM [1,4,5].

Some researchers worked on Sleep Apnea (Apnea) classification such as Tanner et al. and Tari et al. Tanner's team used Decision tree approach and they classified 1200 sleep disorder and found 6 notable features. They got 84% accurateness [6]. Trig's team used Self Organizing MAP (SMO) and ML feed-forward neural networks (MFNN). They clustered sleep disorder into two sets and got only 70% correctness [7]. Derange et al. suggested using decision tree method to classify Sleep Apnea sleep disorder from two data sets [9]. They got 97.6% and 96.6% accuracy from first and second experiment correspondingly. So, experimental results shown that decision tree approach did not counterpart this task very much Wijeya Farooqi et al. categorized Sleep Apnea by using one of classification technique Decision Tree [3]. They used Data Mining techniques for the efficient classification of the Sleep Apnea case. They performed two experimental using Decision tree. The Second experiment classifies Sleep Apnea on the base of expert weighted attributes, which are used in classification on the base of Minimum Cost and source availability. Correctness of this model's still high 98.62%. We coordinated recital in term of Type II error.

M Naresh Kumar used unfit the Decision Tree Approach for the initial diagnosis of Sleep Apnea and accorded its performance with the C4.5 algorithm [10]. A discontinuous Decision Tree technique was able to distinguish the Sleep Apnea using the experimental and laboratory data with the number of correctly classified occurrences as F-measure, and (ROC) as compared to C4.5, h F-measure. Alternating Decision tree-based approach with enhancing has been able to predict Sleep Apnea with a better degree of correctness than the C4.5 based Decision tree using simple medical and workshop features.

Noor Diana et al. presented the Malaysian Sleep Apnea outbreak detection model using three classification methods

[11]. They presented a collection of dissimilar Sleep Apnea data attributes that are used for classification modeling and performances are matched with previous related work. A significant selection of attributes in the Sleep Apnea dataset supports to good results. The Decision tree and Nearest Neighbor models have generally used methods in this problem, while RS was a rule-based method which provides significant knowledge to be further well-thought-out by professionals.

JAVA with WEKA data mining tool was used by Ka shish Ara et al. for Sleep Apnea Disease prediction. Sleep Apnea data was firstly classified and then equated the different Data Mining techniques in JAVA with WEKA through different interfaces as mentioned in (Figure 1) [12]. For categorizing data and to support manipulators in mining useful information from data and readily recognize a suitable technique for precision of analytical typical from that, as it was core objective of their research. The deduction is that NB and J48 are efficient techniques for exactness as less time was consumed for constructing this model through JAVA with WEKA submissions results and they achieved maximum accuracy=100% with 99 correctly categorized instances, maximum ROC=1, had minimum mean absolute error[12].

II. OBJECTIVE

The overall objective of this research is to use rare of the classification methods to determine the population of Sleep Apnea infected cases in UP district and in surrounding areas geographically. So, that we can relate presentation of different organization systems. Unbiased of this study also includes the evaluation of different organization algorithms with the help of graphs, based on our dataset. We have applied all the techniques by using JAVA with WEKA tool and all the procedure of implementation is within it.

III. METHODOLOGY

We used JAVA with WEKA as the DM tool for testing and execution. JAVA with WEKA is a popular set for machine learning software carved in JAVA developed at the University of Waikato, New Zealand [13]. Our focus is on Sleep Apnea testing that whether a patient is affected by Sleep Apnea or not by using some attributes. Based on results, we will show the accuracy of classification techniques and then compare them. It is a very virtuous Data Mining tool for the classification of accurateness, by using the different techniques.

A. Classification

Classification is the case of Data mining, which deals with the tricky things by identifying and noticing features of infection, among sleep disorder and forecast that which technique shows top performance, on the base of JAVA with WEKA's outcome. Five techniques have been used in this work. These techniques use Explorer interface and it depends on dissimilar techniques NB, REP Tree, RT, J48 and SMO. All techniques, which we used, were applied on a Dataset of Sleep Apnea, as enlightened above.

B. DATASET

The Dataset is a collection of data. In the dataset, Most usually a data set resembles to the insides of a single database table, or a single arithmetical data matrix, where every column of the table represents a variable, and each row resembles to a given member of the data set in question. This dataset was taken from District Headquarter Hospital UP. Chunk was selected from this dataset which was treated as Training set and tested this dataset on JAVA with WEKA Data Mining tool. Some data were classified, and rest was tested to check accuracy of data.

C. Attributes

CSV is the file format of datasets which is taken by JAVA with WEKA tool. The Attributes that we have chosen for the challenging of Sleep Apnea are Apnea, Nose bleeding, duration of sleep, Heartbeat rate, fatigue and other indications with class tag of results with positive and negative significances (Figure 2). The attributes description is given in (Table 2).

D. Data Mining Techniques

Different DM techniques have been used for expecting Sleep Apnea virus. These estimates have been done for the purpose of classification and accuracy by using different systems. The edge used for this objective in work is Explorer Interface. Accuracy can be experimental by selecting the following measures: NB, REP tree, RT, J48 and SMO.
The methods we are using are following:

- NB
- REPTree
- RT
- J48
- SOM

Naive Bayes technique: It achieves arithmetical prediction, i.e., estimates class affiliation possibilities. It is based on Bayes formula. A simple NB classifier; confirms equivalent performance with ID3 and selected neural system classifiers. We verified our training set on JAVA with WEKA Data Mining tool with NB Technique, we got the outcomes

mentioned in the (Table 3).

REP tree: Rep Tree uses a regression tree reason and creates several trees in different re-iterations. After that, it picks the best one from all produced trees. That will be measured as the illustrative (Figure 3)

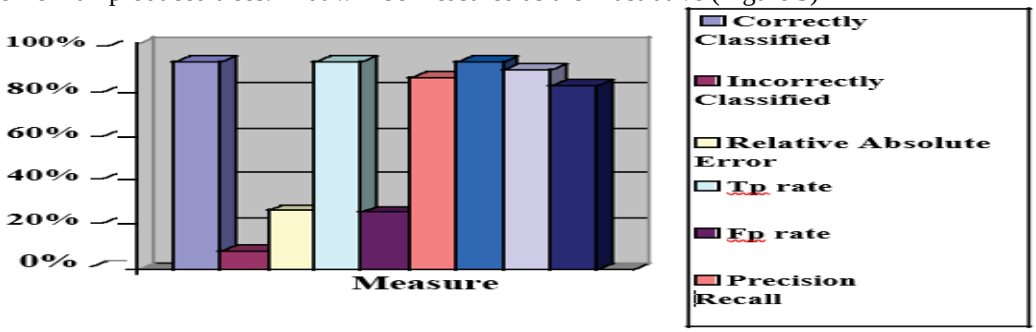

Figure 2: Bayesian graph.

estimations made by the tree. We verified our training set on JAVA with WEKA Data Mining tool with REP tree technique; we got the outcomes mentioned in the (Table 4).

RT: Random Tree is the managed Classifier; it was a collective learning technique which produces many single learners. It employs a transmittable idea to create a set of arbitrary data for building an ID3 (Figure 4). In standard tree near each node is separated using the best split between all variables. In the random forest, every node is split using a best among the subset of predicators arbitrarily chosen at that node [14]. We tested our training set on JAVA with WEKA Data Mining tool with Random tree technique; we got the outcomes mentioned in the (Table5).

Table 2: Attribute description

Attributes	**Description**
Ep-id	id of Patient
Apnea	Yes or no
Nose Bleeding	Yes or no
Flu	Yes or no
Duration of sleep	Yes or no
Others	Other symptoms
Results	Positive or negative

Table 3: Bayesian technique

Attributes name	**Measure**
Correctly Classified	92%
Incorrectly Classified	8%
Relative Absolute Error	26%
Tp rate	0.92
Fp rate	0.253
Precision	0.848
Recall	0.92
F-measure	0.882
Roc Area	0.815

E. J48

C4.5 is the technique used to create a decision ID3 established by Ross Quinlan. C4.5 is an adding of Quinlan's earlier ID3 Method.

The decision trees created by the C4.5 can also be used for organization, and for this purpose, the C4.5 is often specified to as an arithmetical classifier. C4.5 constructs decision trees from the set of exercise data in the identical way as ID3, with the concept of the information entropy (Figure 5). We confirmed our training set on JAVA with WEKA Data mining tool with J48 Technique; we got the results mentioned in the (Table 6).

SMO: SMO is abbreviation of Sequential minimal optimization, which is a method for answering the QP problem that increases during the training of SVM. SMO is extensively used for the training of SVM [15]. We are using this technique on the base of dataset, for intense our data (Figure 6). After running this technique, we assessed the output of classifier by altered measurements to create prediction for each amount of Sleep Apnea dataset. We verified our training set on JAVA with WEKA Data Mining tool with SMO technique; we got the outcomes mentioned in the (Table 7).

F. Comparison

With 5 methods of Data Mining, we have completed classification on our dataset. After examination of our dataset with each technique we are paralleling them in the conclusion. When we have done the comparison among all of them, we decided that the naïve Bayes Technique is extreme among all others. As the accuracy of Naive Bayes is 92% which was major of all. Naïve Bayes is the best also for the aim that it gives the possibility and efficiency while Arbitrary Tree and REP Tree don't give us probability. The below gives the comparison of all the techniques (Table 8). The graph comparison is given in (Figure 7).

IV. CONCLUSION

The main Objective of this work is toward prediction of Sleep Apnea infection using JAVA with WEKA Data Mining tool. Basically, it has four edges. Out of these four edges we are consuming only one edge which is Explorer. We are using five systems of organization, i.e., NB, SMO, J48, RT and REP tree. These techniques were applied using JAVA with WEKA Data Mining tool to evaluate the accuracy which was gained after analysis of these techniques. After challenging these techniques, the result was compared on the base of accuracy. These techniques contest classifier accuracy with each other on base of appropriately classified instances, a precision, error rate, TP rate, FP rate and ROC Area. Over Explorer system it has determined that NB and J48 are the top performance classifier techniques by way that, they has achieved an accuracy of 92% and 88%, takes less time to run and shows ROC area=0.815, and had least error rate.

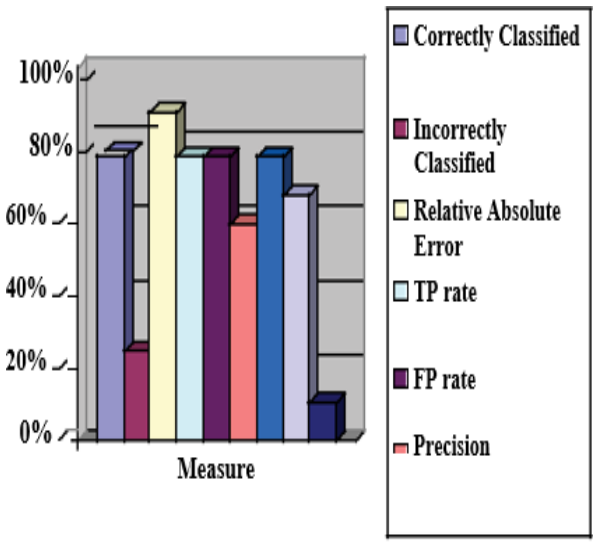

Figure 3: REP graph.

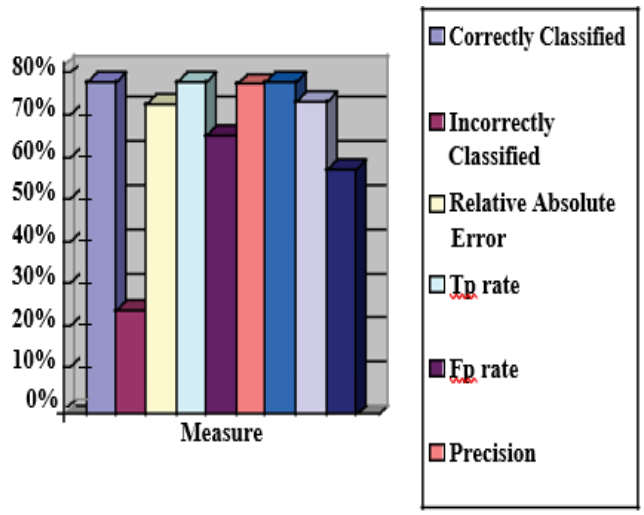

Figure 4: Random forest graph.

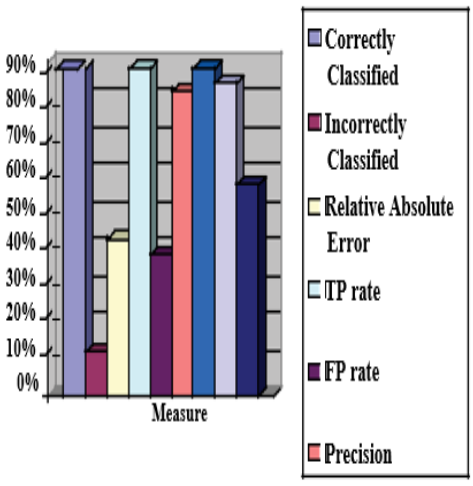

Figure 5: J48 graph.

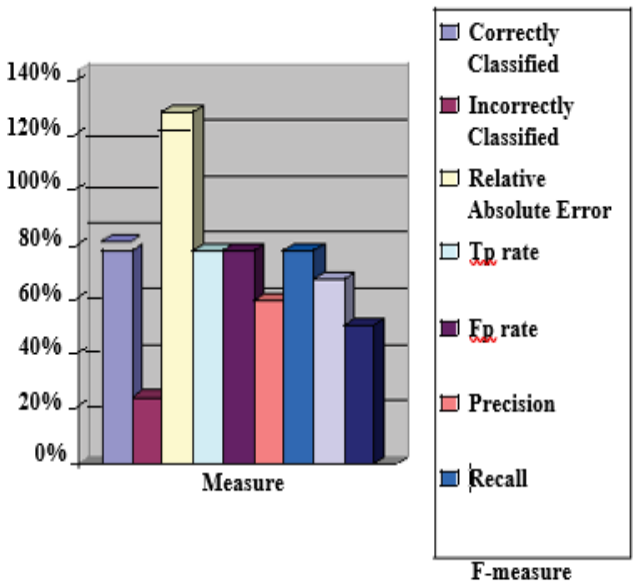

Figure 6: SMO graph.

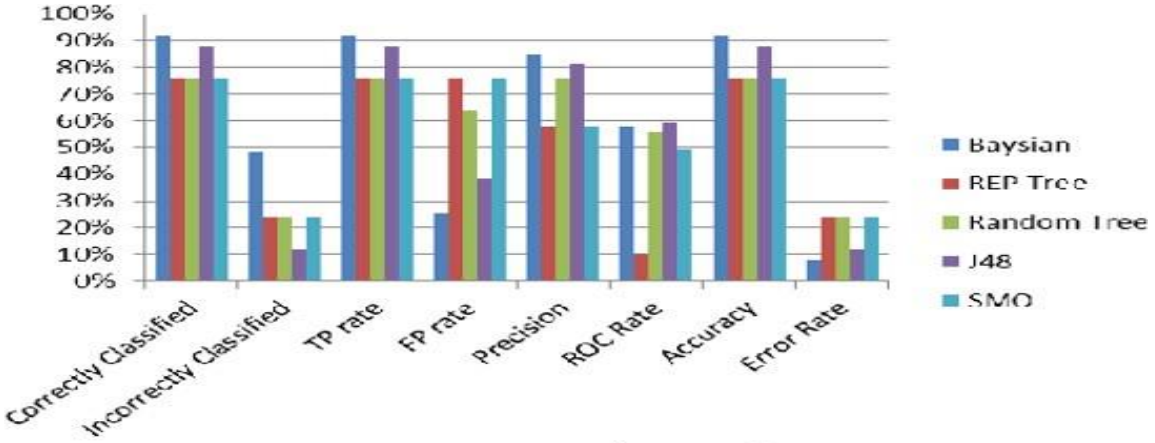

Figure 7: Comparison graph.

Table 4:REP technique

Attributes name	Measure
Correctly Classified	76%
Incorrectly Classified	24%
Relative Absolute Error	88%
TP rate	0.76
FP rate	0.76
Precision	0.578
Recall	0.76
F-measure	0.656
Roc Area	0.099

Table 5: Randomtree

Attributes name	Measure
Correctly Classified	76%
Incorrectly Classified	24%
Relative Absolute Error	71%
Fp rate	0.635
Precision	0.755
Recall	0.76
F-measure	0.715
Roc Area	0.561

Table 6: J48technique.

Attributes name	Measure
Correctly Classified	88%
Incorrectly Classified	12%
Relative Absolute Error	42%
TP rate	0.88
FP rate	0.38
Precision	0.816
Recall	0.88
F-measure	0.842
Roc Area	0.569

Table 7: SMO technique

Attributes name	Measure
Correctly Classified	76%
Incorrectly Classified	24%
Relative Absolute Error	125%
Tp rate	0.76
Fp rate	0.76
Precision	0.578
Recall	0.76
F-measure	0.656
Roc Area	0.494

REFERENCES

[1] Farooqi W, Ali S (2013) A Critical Study of Selected Classification Algorithms for Sleep Apnea and Sleep Apnea Hemorrhagic Apnea. Frontiers of Information Technology (FIT), 11th International Conference on IEEE.
[2] Farooqi W, Ali S, Abdul W (2014) Classification of Sleep Apnea Using Decision Tree. VAWKUM Transaction on Computer Sciences 3: 15- 22.
[3] Rigau-Pérez JG, et.al. (1998) Sleep Apnea and Sleep Apnea Hemorrhagic Apnea. The Lancet 19:971-977
[4] Tanner L, Schreiber M, Low JG, Ong A, Tolfvenstam T, et.al. (2008) Decision Tree Algorithms Predict the Diagnosis and Outcome of Sleep Apnea in the Early Phase of Illness. PLoS Neglected Tropical Disease 12:e196.
[5] Phyu TN (2009) Survey of classification techniques in data mining. Proceedings of the International Multi conference of Engineers and Computer Sci-entists Vol.1.
[6] Vong S, et.al. (2010) Sleep Apnea incidence in urban and rural Cambodia: results from population-based active Apnea surveillance, 2006–2008. PLoS neglected tropical diseases 4:e903.
[7] Faisal T, Ibrahim F, Taib MN (2010) A noninvasive intelligent approach for predicting the risk in Sleep Apnea sleep disorder. Expert Systems with Application 37:2175-2181.
[8] Ibrahim F, Taib MN, Abas WA, Guan CC, Suleiman S (2005) A novel Sleep Apnea (DF) and Sleep Apnea hemorrhagic Apnea (SP) analysis using artificial neural network (ANN). Computer Methods and Programs in Biomedicine 79:273-281.
[9] Daranee T, Prapat S, Nuanwan S (2012) Data mining of Sleep Apnea infection using decision tree. Entropy 2:2.
[10] Kumar MN (2013) Alternating Decision trees for early diagnosis of Sleep Apnea . arXiv preprint arXiv:1305.7331.
[11] Tarmizi NDA, et.al. (2013) Classification of Sleep Apnea Outbreak Using Data Mining Models. Research Notes in Information Science 12:71-75.
[12] Shakil KA,AnisS, AlamM(2015)Sleep Apnea disease prediction using JAVA with WEKA data miningtool.arXivpreprintarXiv:1502.05167.
[13] Pérez MS, et.al. (2005) Adapting the JAVA with WEKA data mining toolkit to a grid based environment. Advances in Web Intelligence Springer, Ber- lin, Heidelberg,492-497.
[14] Gislason PO, Benediktsson JA, Sveinsson JR (2004) Random forest classification of multisource remote sensing and geographic data. Geoscience and Remote Sensing Symposium 2004 IGARSS‘04 Proceedings 2004 IEEE International Vol.2.
[15] Keerthi SS, et.al. (2001) Improvements to Platt‘s SMO algorithm for SVM classifier design. Neural Computation vol.13, pp:637-649.