ISSN 2563-7568

RESEARCH ARTICLE

Development and Investigation of Cost-Sensitive Pruned Decision Tree Model for Improved Schizophrenia Diagnosis

Ephraim Nwoye[1*], Wai Lok Woo[2], Obinna Fidelis[3], Charles Umeh[4], Bin Gao[5]

[1]*Department of Biomedical Engineering, University of Lagos, Lagos, Nigeria*

[2]*Department of Computer and Information Sciences, Northumbria University, England, United Kingdom*

[3]*Department of Biomedical Technology, Federal University of Technology, Akure, Ondo State, Nigeria*

[4]*Department of Psychiatry, Lagos University Teaching Hospital, Idi-Araba, Lagos State, Nigeria*

[5]*School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, China*

Abstract

Schizophrenia is often characterized by delusions, hallucinations, and other cognitive difficulties; affects approximately seventy million adults globally. This study presents a cost-sensitive pruned Decision Tree J48 model for fast and accurate diagnosis of Schizophrenia. The model implements supervised learning procedures with 10-fold cross-validation resampling method and utilizes unstructured filter to replace missing values in the data with the modal values of corresponding features. Features are selected using Pearson's correlation on hot-coded data to detect redundancy in data. Cost matrix is designed to minimize the tendencies of the J48 algorithm to predict false negative outcomes. This consequently reduces the error of the model in diagnosing a Schizophrenia candidate as free from the disease. The model is found to significantly diagnose Schizophrenia with 78% accuracy, 89.7% sensitivity, 57.4% specificity and Area under the Receiver Operator Characteristic (ROC) curve of 0.895. The ROC curve is also seen to distinguish Schizophrenia from other conditions with similar symptoms. These results show the potential of machine-learning models for quick, effective diagnosis of schizophrenia.

Key Words: *Schizophrenia; Decision tree; Machine learning; Naive-Bayes model*

***Corresponding Author**: Ephraim Nwoye, Department of Biomedical Engineering, University of Lagos, Lagos, Nigeria, Tel: +234 8033735926; E-mail: enwoye@unilag.edu.ng*

***Received Date**: July 20, 2020, **Accepted Date**: September 07, 2020, **Published Date**: October 30, 2020*

***Citation**: Nwoye E, Woo WL, Fidelis O, et al. Development and Investigation of Cost-Sensitive Pruned Decision Tree Model for Improved Schizophrenia Diagnosis. Int J Auto AI Mach Learn. 2020;1(1):17-41.*

OPEN ACCESS

This open-access article is distributed under the terms of the Creative Commons Attribution Non-Commercial License (CC BY-NC) (http://creativecommons.org/licenses/by-nc/4.0/), which permits reuse, distribution and reproduction of the article, provided that the original work is properly cited and the reuse is restricted to non-commercial purposes.

1. Introduction

Schizophrenia, often characterized by delusions, hallucinations, and other cognitive difficulties, affects about 1 per cent of the global population [1]. The diagnosis is complex, and it takes several processes to arrive at the right decision. Psychiatrists have to watch that the observed symptoms are persistent for at least a month as contained in DSM-V classification details and that those symptoms are not as a result of other related diseases nor a behavioural activity (e.g. alcohol consumption) nor a result of a pharmacological side effects. These requirements have made the process of diagnosis lengthy and complex. Psychiatrists need a quicker, effective, and knowledge-based technique to diagnose Schizophrenia for early detection, prompt treatment and effective management of the disease.

Data mining is a discipline which is driven majorly by the growth of large databases that contain large information which are concealed within a mass of uninteresting data but the interesting salient features can be revealed. In other words, it is the discovery of knowledge from data and consists of data cleaning, integration, selection, transformation, mining, pattern evaluation and knowledge presentation [2]. The subject can be described as knowledge discovery from data and is closely allied to exploratory data analysis. Many tools are used in data mining. Some recent examples in pattern recognition and detection include tools for characterizing, identifying, and locating patterns in multivariate response data as well as tools for detecting and identifying patterns in two-dimensional displays. Examples of important tools in model building in data mining include recursive partitioning, cluster analysis, regression modeling, segmentation of time series into a small number of segment types, techniques for condensing huge data sets into manageable summaries, and collaborative filtering, in which transactions are processed as they arrive so that future transactions may be treated in a more appropriate manner [3].

Important information can be obtained from health records of individuals using data mining techniques [4] and can be used to predict the likelihood that a person will suffer Schizophrenia. This kind of systems can serve as an adjunct tool for psychiatrists and medical students to diagnose patients with Schizophrenia. Data mining involves the use of algorithms to find relationships among features defining a dataset. These algorithms transform information into actionable intelligence. Examples of such algorithms include Decision Trees, Bayesians, Regressions, Neural Networks, Support Vector Machines (SVM), KNN, etc. Many of these algorithms are available in data mining software such as R, Python, Java, WEKA, etc.

The cost-sensitive J48 model has been used for a number of predictive and analytical studies and in various applications such as prediction of soil fertility [5], assessment of solar potential of western Himalayan Indian state of Himachal Pradesh [6], intrusion detection system [7] and for predicting the incidence of diabetes [8] amongst other applications. In one study [9], the model was used to predict the incidence of myocardial infarction in humans. The study conducted in Iran involved 750 patients (455 healthy and 295 myocardial cases) and utilized 92 regular features (such as demographics, blood pressure, body mass index, etc.) and 1 label feature. The data was pre-processed and 90% of the data was used for training the model and 10% was used for testing. The model achieved 86.67% sensitivity, 80% F-measure and 82.67% accuracy.

In another study [10], researchers extracted EEG signals with 21 gold cup electrodes placed according to the 10-20 international system while monitoring the horizontal and vertical eye movements of participants. Some software such as MATLAB and EEG Lab toolboxes were used in

the signal preprocessing and five frequency bands were selected for analysis. The receiver operating curve (ROC) was done to differentiate between normal and schizophrenia patients. It was reported that the delta power had the highest classification accuracy of 62.2%. The study further researched EEG recordings on 50 participants using 64 electrodes. The electrodes were placed above and beneath the right eye and laterally concerning the right and left eyes to monitor horizontal and vertical movements, respectively. The extracted features were used to represent normal and schizophrenia patients. This produced a better classification accuracy with five different electrodes having a prediction accuracy ranging from 92.0% and 93.9%.

In another study [11], Schizophrenia was classified using MR imaging data in two fields (3- and 7-T). Gray matters and White matters were used as inputs into a Support Vector Machine (SVM). The 7-Tesla classifiers outperformed the 3-T classifiers by as much as 77% compared to 66.6% accuracy for the 3-T classifier. The study in [1] proposed a SVM model for diagnostic classification of schizophrenia patients on the basis of Regional Reward-related fMRI signal patterns. The study investigated whether the predictive accuracy for the diagnostic classification of schizophrenia patients vs. healthy controls could be improved using multivariate pattern analysis (MVPA) of regional functional magnetic resonance imaging (fMRI) activation patterns for the anticipation of monetary reward. The results show that MVPA can be used to substantially improve the accuracy of diagnostic classification on the basis of task-related fMRI signal patterns in a regionally specific way.

In [12], a neural network model was developed to predict the likelihood of developing number of psychological conditions such as anxiety, behavioral disorders, depression, and posttraumatic stress disorder. The study tested the effectiveness of the model against a dataset of 89,840 patients. The results show that the model is capable of achieving accuracies ranging from 73% to 95% for each of the clinical conditions under consideration, with an overall accuracy of 82.35% for all conditions.

The study [13] involved a machine-learning-based diagnosis of Schizophrenia using combined sensor-level and source-level EEG levels. Sensor-level and source-level features were extracted from EEG signals recorded during an auditory oddball task for the classification of patients with schizophrenia and healthy subjects. The selected sensor-level features were mostly found in the frontal area, and the selected source-level features were mostly extracted from the temporal area, which coincide well with the well-known pathological region of cognitive processing in patients with schizophrenia. In [14], deep-learning method for recognition of early-onset schizophrenia was deployed. A three-stage deep-learning network was used to deduce dimension reduction, and feed-forward back propagation neural networks were used as classifiers. The classification accuracy reached 79.3% with 87.4% sensitivity and 82.2% for specificity. The study showed that resting-state connectivity presented good potential classification capacity and can be used as a biomarker of clinical diagnosis of schizophrenia. The monotonous nature of schizophrenia diagnosis coupled with the similarity of symptoms with those of other conditions make diagnosis difficult and time-consuming for psychiatrists. A recent study [15] made electroencephalography recordings of schizophrenia patients and controls viewing natural scenes. They were able to use a rule-based classifier to discriminate schizophrenia patients from controls and obtained an accuracy of 71%; specificity of 81% and sensitivity of 59% (Table 1).

Table 1:
Summary of results from reviewed studies.

Study	Results as reported
[9]	The model achieved 86.67% sensitivity, 80% F-measure and 82.67% accuracy
[10]	An 88.53% accuracy and F1 score of 91.22% was obtained
[11]	77% compared to 66.6% accuracy for 3-T and 7-T classifiers
[12]	overall accuracy of 82.35%
[14]	Accuracy reached 79.3% with 87.4% sensitivity and 82.2% for specificity
[15]	An accuracy of 71%; specificity of 81% and sensitivity of 59%

This paper aims to develop a machine learning algorithm for the quick diagnosis of schizophrenia. The main contribution of this study is a novel cost-sensitive pruned decision tree J48 model that automates diagnosis of Schizophrenia. The model also minimizes, by penalizing through cost matrix design, the error of making costlier decision errors that can result in diagnosing other diseases instead of Schizophrenia. This work also contributes an optimal treatment regime workflow that can assist psychiatrists in effective decision making in the management process of Schizophrenia.

2. Methodology

2.1. Data Acquisition and Preparation

Data used for the present study were collected from the Psychiatric Clinic of the Lagos University Teaching Hospital, Lagos, Nigeria. Dataset consists of 151 health records of patients reported between years 2013 and 2018 inclusive. 105 records are of positively diagnosed Schizophrenia patients, and 46 are of those diagnosed otherwise, to serve as controls in the study. Thirty-three (33) attributes based on Diagnostic and Statistical Manual of Mental Disorders (DSM-5) specifications for definition of Schizophrenia were used in model building with CLASS as labels excluding Year of Patient Report, Age, Sex, and Diagnosis. This is shown in Appendix 1. The unstructured file record dataset was converted into electronic structured data using Ms Excel and saved as '.csv' file acceptable in Waikato Environment for Knowledge Analysis (WEKA) API software [16]. The data was loaded on Python environment for observation and visualization. Missing values were replaced with modal values of the corresponding features. The flowchart for the development of the model is shown in Figure 1.

2.2. Cost Sensitive Pruned J48 Classifier

Decision trees are learned from training data. Each data item consists of a set of features describing an object and the class of the object. Decision trees are recursively built beginning with the topmost node by:

i Computing the best test for the current node according to some splitting criterion.
ii Creating a sub-node for each possible outcome of the test.

iii Recursively expanding each sub-node in the same way until a given stopping criterion is satisfied.

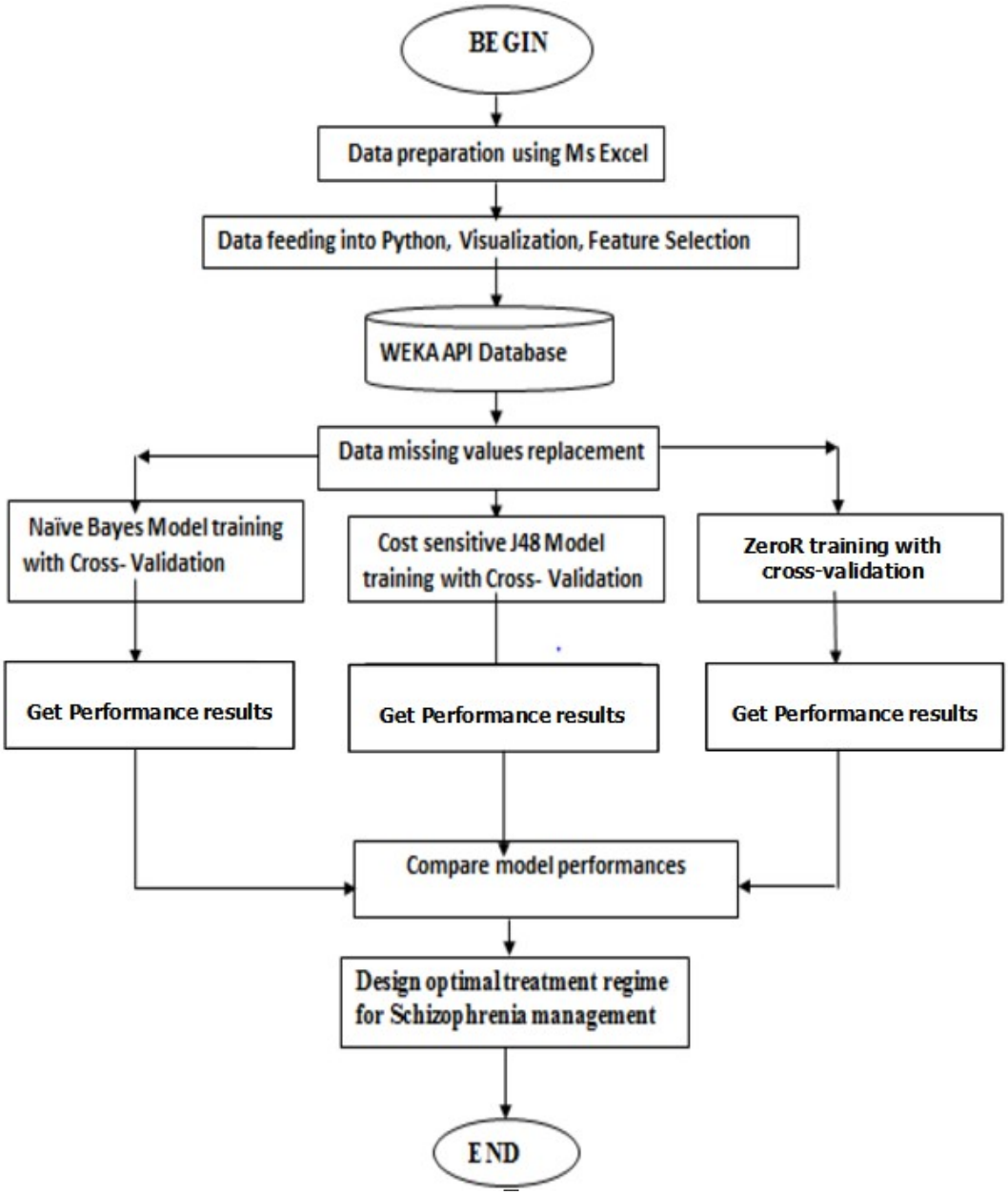

Figure 1: *Flowchart of the Methodology*.

Usually, the decision tree is afterwards simplified (pruned) in order to avoid over-fitting of the training data. The test of the topmost node divides the training data into two subsets. One subset contains the elements which pass the test; the other contains the elements which fail the test. In order to avoid over-fitting, most decision tree learning algorithms add another step which simplifies the decision tree with pruning. Pruning identifies irrelevant feature tests and replaces the corresponding non-terminal nodes with terminal nodes.

Cost sensitive: Model's prediction is optimized by making mistakes costlier than others. Diagnosing a patient negative when actually the patient is positive can be an expensive mistake. One solution is to minimize the number of false negatives predictions. J48 algorithm allows us to

assign different penalty to different types of errors, in order to discourage a tree from making more costly mistakes. The penalty is designated in a cost matrix, which specifies how much costlier each error is relative to any other predictions. In this work, it assumes that false negative diagnosis costs the patient two times as much as false positive diagnosis. The algorithm awards no cost when an instance is classified correctly, but a false negative prediction attracts a penalty cost of 2 versus a false positive's cost of 1. For an ordinary decision tree model, no such penalty is levied on the false predictions.

Splitting criterion: The best test for a node is selected according to the splitting criterion. A frequently used splitting criterion is the information gain. It is the difference between the entropy of the data set at the current node and the entropy in the two subsets induced by the test. The entropy of a data set measures to which degree the data is scattered over several classes. If a data set is pure, i.e., if all elements belong to the same class, the entropy is 0. If half of the data belongs to class A and half of the data to class B, the entropy is 1. The entropy is defined by Equation (1):

$$H(p)=-\sum_{c\in C} p(c)\log_2 p(c) \tag{1}$$

where $p(c)$ is the relative frequency (empirical probability) of class $c\in C$ in the data set, i.e., the frequency of class c divided by the size of the data set. The information gain is defined by Equation (2):

$$G=H(p)-w_1 H(p_1)-w_2 H(p_2) \tag{2}$$

where p is the relative frequency of class c in the current data set, p_1 and p_2 are the relative frequencies of class c in the two subsets, and w_1 and w_2 are the proportions of data in the first and second subset.

Stopping criterion: The recursive expansion of the decision tree is stopped if either the data subset is pure or if all data items have the same feature representation. The latter case occurs when the data contains objects with identical feature values, but different classes. When such contradictory class assignments exist, there is no decision tree which correctly classifies all the training data. Sometimes these stopping criteria are augmented by other criteria which may terminate the induction process earlier, such as:

(i) The size of the data set being below a certain threshold
(ii) The value of the splitting criterion for the best test being below a threshold

Pruning: Decision trees which are grown to their maximal size as described tend to over-fit the training data. Overfitting occurs when the classification of the decision tree depends on accidental properties of the training data. Overfitting is a problem because it leads to errors on new data. In order to avoid overfitting, most decision tree learning algorithms add another step which simplifies the decision tree with pruning. Pruning identifies irrelevant features and replaces the corresponding non-terminal nodes with terminal nodes. The pruned trees are evaluated on test data by computing the classification accuracy, which is the proportion of correctly classified test items. The tree with the highest accuracy is selected. It is important that the data used for this

evaluation is fresh data which was not used to induce the tree. Otherwise, the tree will not be pruned because the full tree achieves the highest accuracy on the training data.

2.3. ZeroR Classifier

A baseline classification uses a naive classification rule such as:

(i) Base Rate: Accuracy of trivially predicting the most frequent class. ZeroR classifier in WEKA always classify to the largest class i.e. according to the prior.

(ii) Random Rate: Accuracy of making a random class assignment. It might apply prior knowledge to assign random distribution.

(iii) Naive Rate: Accuracy of some simple default or pre-existing model.

ZeroR gives a baseline accuracy that must be always checked before choosing a sophisticated classifier. Its Accuracy is also known as null rate.

2.4. Naive Bayes Classifier

It is a classifier that works based on Bayes Theorem represented by equation (3).

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)} \tag{3}$$

Naive Bayes classifier calculates the probabilities for every factor and then it selects the outcome with highest probability. The classifier assumes the features are independent and hence the word Naive. It is a powerful algorithm used for many classification problems including medical disease diagnosis.

3. Data Preparation and Models Development

The unstructured file record dataset is converted into electronic structured data using Ms Excel and saved as .csv file acceptable in Waikato Environment for Knowledge Analysis (WEKA) API software. The data was loaded on Python environment for observation and visualization. The description of the cleaned dataset is given in Table 2. It is seen clearly that there are missing values in the data as number of entries against some of the features is less than 151.

Figure 2 shows the heatmap of missing value positions in white stripes in dataset. These missing values have been replaced the mean values in corresponding features. This is necessary in order not to miss a lot of information from the data.

Table 2:
Dataset Description.

RangeIndex:	151 entries, 0 to 150
Data columns (total 34 columns):	OCCUP 145 non-null object MAR_STA 149 non-null object DUR_EPIS 148 non-null float64 P_PSY_HX 134 non-null object P_MED_HX 121 non-null object FAM_P_HX 130 non-null object P_SOC_HX 138 non-null object P_SEX_HX 123 non-null object FOR_HX 131 non-null object PREMOBD_HX 123 non-null object MSE 151 non-null object SPEECH 144 non-null object MOOD 143 non-null object AFFECT 141 non-null object TH_FORM 108 non-null object TH_STRM 112 non-null object TH_CONTENT 124 non-null object TH_POSS 93 non-null object PERCEP 112 non-null object ORIENT 120 non-null object ATTEN 126 non-null object CONC 126 non-null object MEM_IR 116 non-null object MEM_ST 116 non-null object MEM_LT 116 non-null object INT_GFK 89 non-null object INT_S_A_D 87 non-null object INT_CAL 84 non-null object INT_PROV 84 non-null object JUDGMT 120 non-null object INSIGHT 131 non-null object PSE 147 non-null object EEG 151 non-null object CLASS 151 non-null object
dtypes:	float64(1), object(33)

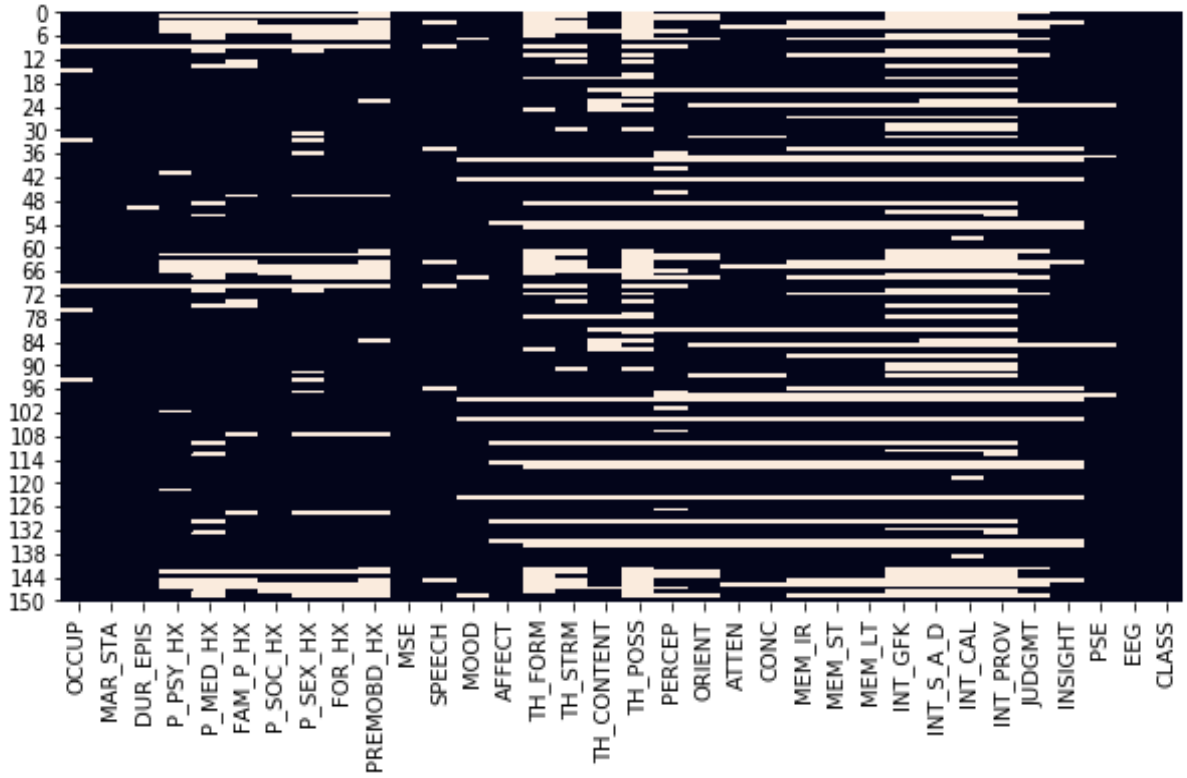

Figure 2: *Heatmap showing missing value positions in white stripes.*

3.1. Feature Selection

In order to clean the dataset of redundant attributes, the dataset was hot-coded and converted to numerical representation. Pearson's correlation was applied to the data to identify fully correlated features. Pearson's correlation coefficient is the covariance of the two variables divided by the product of their standard deviations. The form of the definition involves a "product moment", that is, the mean (the first moment about the origin) of the product of the mean-adjusted random variables; hence the modifier product-moment in the name. The coefficient of correlation when applied to two variables X and Y is given by Equation (4):

$$R_{xy} = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}} \quad (4)$$

where x_i is instant value of X; $\overline{x}$ is mean value; y_i is instant value of Y; and $\overline{y}$ is mean value of Y. Correlation coefficient between two features has value ranges from -1 to +1. The value of -1 represent full negative correlation, 0 means no correlation and +1 stands for full positive correlation.

Figure 3 shows the heatmap of the Pearson's correlation values. From the figure, it is clear that full correlations exist between ATTEN and CONC, among MEM_IR, MEM_ST, MEM_LT, and among INT_GFK, INT_S_A_D, INT_CAL, and INT_PROV. Consequently 'ATTEN', 'MEM_ST', 'MEM_LT', 'INT_S_A_D', 'INT_CAL', 'INT_PROV' features were dropped from the dataset to prevent over-fitting and for effective model performance.

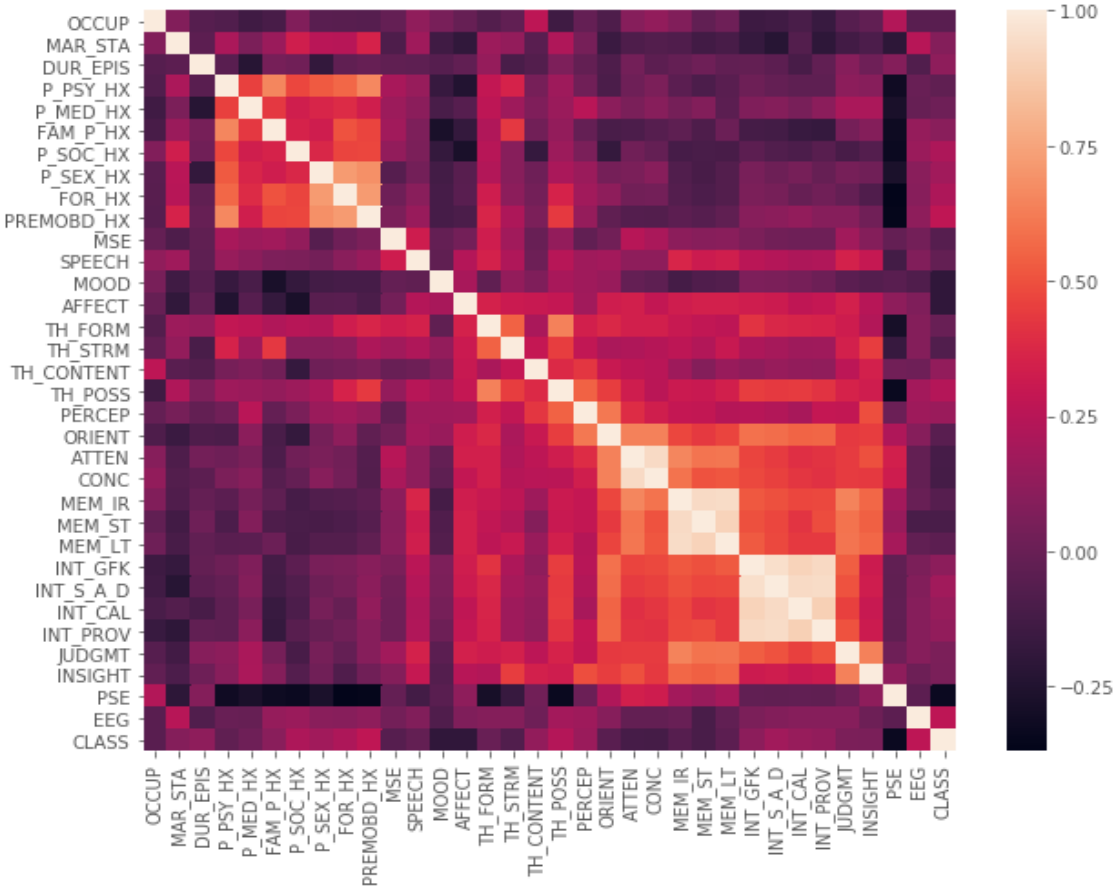

Figure 3: *Pearson's correlation values between each pair of attributes*.

3.2. Models Development

The development of the models involves the following steps:

(i) Data Resampling: Data was divided into training and test datasets. The data was divided into 10 folds. The training data was made 90 per cent of the whole dataset, while the test data constitutes remaining 10 per cent, and training was done repeatedly using cross-validation. This is to estimate the performance of a machine learning algorithm with less variance than a single train-test set split. It works by splitting the dataset into k-parts (e.g. k= 5 or k=10). Each split of the data is called a fold. The algorithm is trained on k−1 folds with one held back and tested on the held back fold. This is repeated so that each fold of the dataset is given a chance to be the held back test set. After running cross-validation, k different performance scores are obtained that can be summarized using mean and standard deviation. The obtained result is a more reliable estimate of the performance of the algorithm on new data. This is because the algorithm is trained and evaluated multiple times on different data. The choice of k must allow the size of each test partition to be large enough to be a reasonable sample of the problem, whilst allowing enough

repetitions of the train-test evaluation of the algorithm to provide a fair estimate of the algorithms performance on unseen data. For modest sized datasets in the thousands or tens of thousands of records, k values of 3, 5 and 10 are common.

(ii) Model Building: The cost-sensitive pruned J48 algorithm is trained using 10-fold cross-validation techniques. Here, we have built the decision-tree models using cost-sensitive pruned J48 algorithm and ordinary J48 algorithm with prune-only functionality. The trained models are shown in Appendix 2 and Appendix 3, respectively. The resulting decision-tree built using the cost-sensitive pruned J48 algorithm has a size of 36 trees and 30 leaves while the decision-tree built using the ordinary J48 algorithm has a size of 41 trees and 33 leaves. A Dell Inspiron 11 3000, 8 GB RAM, 500 GB Hard Disk Laptop Computer was used for the development. The implementation of the proposed framework was performed in Python distributed open source by the Anaconda Integrated Development Environment and included the use of in-built packages such as Keras, Theano, Scikilearn, Matplotlib, Seaborn, Numpy.

(iii) Model Evaluation: The test data folds are used to evaluate the predictive performance of the model.

(iv) Optimization of Model's Prediction: Model's prediction was optimized by making mistakes costlier than others. Diagnosing a patient negative when he is positive can be an expensive mistake. One solution is to minimize the number of False Negatives predictions. J48 algorithm allows us to assign different penalty to different types of errors, in order to discourage a tree from making more costly mistakes. The penalties are designated in a cost matrix, which specifies how much costlier each error is relative to any other predictions. It was assumed that false negative diagnosis costs the patient two times as much as false positive diagnosis. Algorithm awards no cost when an instance is classified correctly, but a false negative prediction attracts a penalty cost of 2 versus a false positive's cost of 1.

(v) Building Baseline Model, ZeroR: The same datasets i.e., training and test-are used to develop the baseline model ZeroR. ZeroR is the most rudiment of classifiers. It simply predicts the majority if data is nominal, and the mean value if data is numeric. It may overfits depending on skewness of the dataset. However, it is commonly used to validate the performances of other complex classification algorithms on a given dataset, especially if the dataset is skewed. The model predicted 97 instances accurately and 54 wrongly and as a result had an accuracy of 64 per cent. This is a reflection of skewness in dataset used; the dataset has 97 SCHIZ instances and 54 OTHERS instances.

(vi) Building Naive Bayes Model: Naive Bayes classifier predicts the instance X belongs to the class C_i if and only if:

$$P(C_i \vee X) > P(C_j \vee X) \text{ for } 1 \le j \le m, j \ne i$$

where $P(C_i|X) = \frac{P(X \vee C_i) P(Ci)}{P(X)}$ .The Naive Bayes algorithm has been applied to the dataset and had an accuracy of 82 per cent. When cost matrix was added the accuracy improved to 83 percent.

The resulting trained model template is shown in the Appendix 4. The posterior probability $P(C_i \vee X)$ has been computed for each attribute using the Bayes theorem.

(vii) Models Validation: ZeroR is a baseline algorithm commonly used in validating machine learning model performance. Naive Bayes algorithm is a classifier reportedly used by medical scientists in classification problems. Its performances have been adjudged in literatures as reasonable. The two algorithms (ZeroR and Naive Bayes) were also trained on the project datasets. Performance of cost-sensitive J48 in classifying the instances into appropriate classes was compared with performances of the two validation classifiers to establish the reliance of CS J48 in diagnosis of Schizophrenia. The comparison of the Baseliner Classifier, ZeroR, ordinary J48, Naive Bayes, and CS J48 classifiers are done on the basis of Receiver Operating Characteristics (ROC) curve, area under ROC, confusion matrices, sensitivity, specificity, and diagnostic odds ratio.

3.3. Performance Statistic Metrics

Confusion matrix: The confusion matrix of a classifier shows the instances of true positives (TP), false negatives (FN), false positives (FP) and true negatives (TN) in an array. Associated with the confusion matrix is the accuracy which is given by

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

Sensitivity: This measures the proportion of positives that are correctly identified as such. This is an important metric in medical diagnostics and perfect predictor is 100 % sensitive.

$$Sensitivity = \frac{TP}{TP + FN}$$

Specificity: This measures ability of classifier to correctly identify those without Schizophrenia from a dataset. Best specificity is 1 whereas the worst is 0.

$$Specificity = \frac{TN}{TN + FP}$$

Diagnostic Odds Ratio: In medical testing, is a measure of effectiveness of a diagnostic test. It is defined as the ratio of the odds of the test being positive if the subject has a disease relative to the odds of the test being positive if the subject does not have the disease. The rational for the Diagnostic Odds Ratio is that it is a single indicator of test performance (like accuracy) but which is independent of prevalence (unlike accuracy) and is presented as an odds ratio, which is familiar to medical practitioners. Mathematically, Diagnostic Odds Ratio (DOR) is defined as:

$$DOR = \frac{\frac{TP}{FP}}{\frac{FN}{TN}} = \frac{TP}{FP} \cdot \frac{TN}{FN}$$

DOR value ranges from zero to infinity, although for useful test it is greater than one, and hvigher diagnostic odds ratios are indicative of better test performance.

ROC Curves: In a Receiver Operating Characteristic (ROC) curve the true positive rate (Sensitivity) is plotted in function of the false positive rate (100-Specificity) for different cut-off points. Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. A test with perfect discrimination (no overlap in the two distributions) has a ROC curve that passes through the upper left corner (100% sensitivity, 100% specificity). The closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test.

Area under ROC Curve (AUC): This is a performance metric for binary classification problems. The AUC represents a model's ability to discriminate between positive and negative classes. An area of 1.0 represents a model that made all predictions perfectly. An area of 0.5 represents a model that is as good as random.

4. Results and Analysis

The training implemented 10-fold cross-validation train and test mode to eliminate bias due to random sampling. The results for the J48 model returned an area under ROC 0.943, and accuracy of 86%. The Confusion matrix is shown in Table 3.

Table 3:
Confusion matrix of pruned J48 model.

.

ACTUAL CLASS	PREDICTED CLASS	
	SCHIZ	OTHERS
SCHIZ	81	16
OTHERS	4	50

The result in Table 3 implies that the J48 model is able to predict 81 cases accurately. The number of patients predicted to have the disease but in actual case do not have it is 4. Those who have schizophrenia but are predicted not to have it are 16 and 50 patients do not have schizophrenia and are so predicted. This interpretation applies to all the confusion matrices in the other tables. CS-J48 gave accuracy of 78% and ROC area 0.895. The confusion matrix of CS-J48 model is shown in Table 4. It can be seen that the resulted confusion matrix shows less FN predictions although accuracy has been traded off.

Table 4:
Confusion matrix of cost-sensitive J48 pruned (CS-J48) model.

ACTUAL CLASS	PREDICTED CLASS	
	SCHIZ	OTHERS
SCHIZ	87	10
OTHERS	23	31

The ZeroR model is developed using the same dataset and training procedures in Section 3. The area under ROC curve is computed to be 0.467 with accuracy of 64%. The Confusion matrix is shown in Table 5.

Table 5:
Confusion matrix of ZeroR model.

ACTUAL CLASS	PREDICTED CLASS	
	SCHIZ	OTHERS
SCHIZ	97	0
OTHERS	54	0

A Naive Bayes model is developed using the same dataset and using the training procedures as described in Section 3. The area under ROC curve is 0.917 and with accuracy of 82%. The Confusion matrix is shown in Table 6.

Table 6:
Confusion matrix of Naive Bayes model.

ACTUAL CLASS	PREDICTED CLASS	
	SCHIZ	OTHERS
SCHIZ	80	17
OTHERS	10	44

4.1. Validation of Model

The cost-sensitive J48 model developed for quick diagnosis of Schizophrenia was internally validated by Cross-Validation test mode. The external validation was done by comparing its (CS J48's) performance against performances of J48, Naive Bayes and ZeroR models developed on the same dataset used in building CS-J48 model for Schizophrenia diagnosis. Table 6 shows the comparisons are carried out on the basis of Area of Receiver Operating Characteristics curve (Area of ROC), accuracy, sensitivity, selectivity, and diagnostic odd ratio. Table 7 and Figures 4 (a-d) show the results of the performance metrics and the ROC curves for each model.

Table 7:
Performance metrics of models.

MODEL	ACCURACY (%)	SENSITIVITY	SPECIFICITY	DOR	ROC AREA
CS-J48	78	0.897	0.574	21	0.895
J48	86	0.835	0.926	63	0.943
NAIVE BAYES	82	0.825	0.814	21	0.917
ZeroR	64	1.000	0	Undefined	0.467

Sensitivity refers to the number of subjects who have schizophrenia and were so predicted. The two best models under sensitivity metric are given by ZeroR and CS-J48. Specificity refers to the number of persons who do not have the condition and were so predicted. The two best models under specificity metric are given by J48 and Naive Bayes. The area under the ROC is indicative of the diagnostic ability of the classifiers. Under this metric, all the three models CS-J48, J488 and Naive Bayes perform much better than the baseline ZeroR. This is also reflected in the accuracy metric where J48 has the highest score, followed by Naive Bayes and CS-J48. In terms of DOR, J48 has the highest value and this is followed equally by CS-J48 and Naive Bayes. The efficiency of a classifier will therefore be a function of all these parameters.

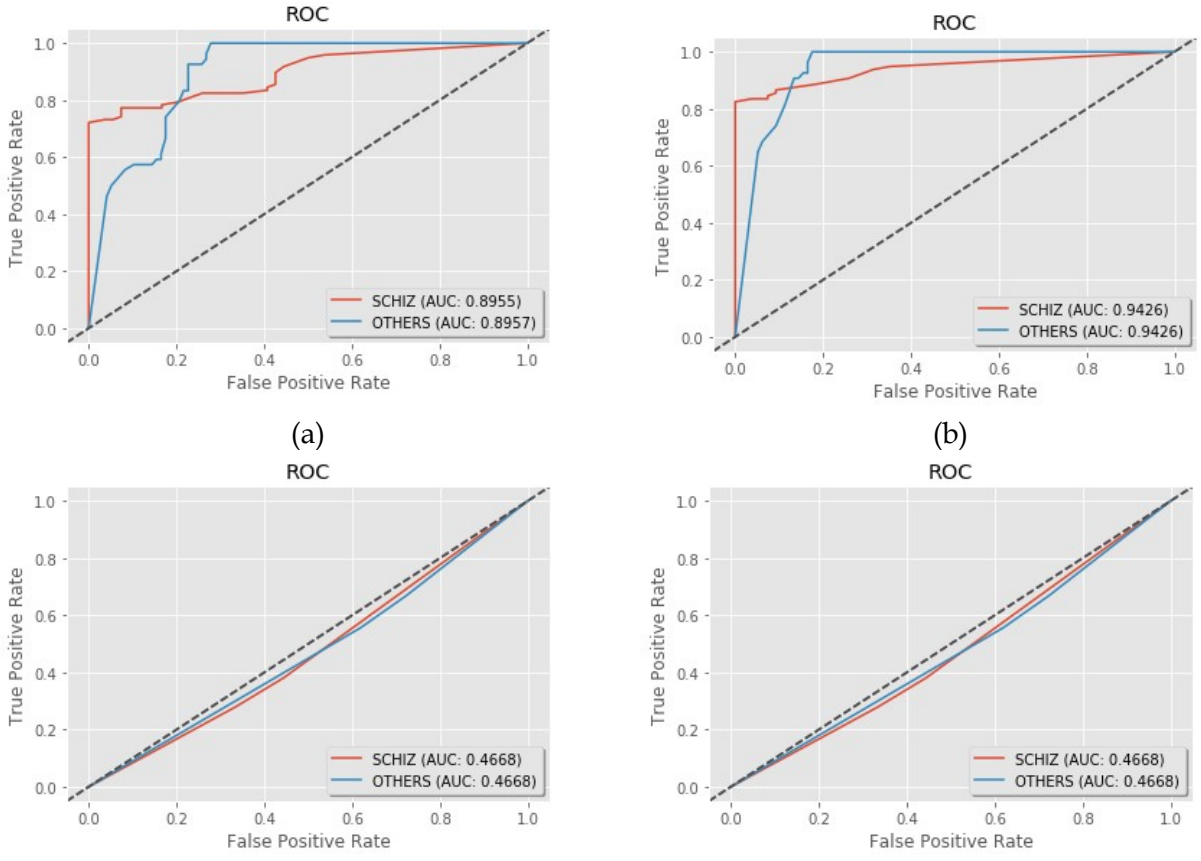

Figure 4: *ROC curves of the models. (a) CS J48 model, (b) J48 model, (c) Naive Bayes model, (d) ZeroR model.*

4.2. Discussion

Performance parameters of the cost-sensitive J48 (CS-J48) model, J48, ZeroR, and Naive Bayes have been compared to establish the effectiveness of CS-J48 model in Schizophrenia diagnosis. The results show the CS-J48 model performed reasonably well as a disease outcome predictor. The CS J48 model's accuracy is encouraging at 78% and can be reliably used to determine whether a new instance of similar features has Schizophrenia or not. The CS-J48 had better sensitivity than Naive Bayes and J48. It means its probability of predicting positive instances as positive is higher than that of Naive Bayes. However, both cannot be used to rule in presence of Schizophrenia as their sensitivities are high. The specificity of CS-J48 is much lower than that of Naive Bayes; this indicates CS-J48 is a predictor that can be used to rule out Schizophrenia anytime the outcome is negative. Also, area under Receiver Operating Characteristics (ROC) curve determines the discriminant power of a model. As shown, the areas under the ROC curve for the J48 and Naive Bayes curve are comparable (0.895 and 0.917 respectively). This means that the two models can discriminate almost equally between positive and negative class instances. The Diagnostic Odd Ratio value is also another performance parameter used in medical field to measure performance of diagnostic tests. It is defined as the ratio of the odds of the test being positive if the subject has a disease relative to the odds of the test being positive if the subject does not have the disease. Good DOR value is greater than 1 and the greater the better. In the present study, CS J48, Naive Bayes and J48 have DOR greater than 1, making CS-J48 a good diagnostic model. For ZeroR, DOR is undefined indicating that it is a biased predictor. The ROC curve shows discriminating power of a model and the ROC curve of the CS-J48 model show discrimination of the two classes SCHIZ and OTHERS and with good accuracy.

It is noted that when comparing the models in terms of the sensitivity metric, CS-J48 comes out as the top in the list with a score close to 0.9. This is not surprising because the cost matrix has been added to the optimization procedure. The CS-J48 model is specially trained to penalize the false negative error twice more heavily than the false positive. Thus, it is much more sensitive to false negative error. On the other hand, J48 model treats both the false negative and false positive with the same importance as true detection (i.e., true positive and true negative). The model awards no additional penalty when an instance is classified incorrectly. Hence, J48 model has a relatively uniform performance in terms of sensitivity and specificity. The CS-J48 model minimizes prediction of false negative with the cost matrix defined for the algorithm. As shown in the confusion matrices in Tables 3 and 4, the error is more shifted to false positive than false negative in CS-J48 model when compared with ordinary J48 model. Similarly, the ROC curve of CS-J48 ROC shows a clear discrimination of the two classes SCHIZ and OTHERS and with good level of accuracy. Thus, if the diagnostic tool is designed with an aim of reducing miss detection i.e., false negative, then the CS-J48 model will render a better performance than J48 and Naïve Bayes models.

In terms of medical applications, the CS-J48 model can be used as a diagnosis model for optimal treatment of first episode Schizophrenia. The model can be adapted for optimization of existing treatment for Schizophrenia designed through European Union sponsored European Union First Episode Schizophrenia Trial [17]. The model may serve as a guide to psychiatrists in decision making while design treatment plans for patients, especially first episode patients. Quick and accurate detection of Schizophrenia ensures favorable treatment and management outcome as duration of psychotic episode is a factor in efficacy of antipsychotics. In addition, such model can

be embedded into wearable watch [18,19] for non-intrusive monitoring of mental health status beyond clinical settings [20-23], or mobile application for public use [24-26].

5. Conclusion

This study presents the development and investigation of a cost-sensitive pruned Decision Tree J48 model that can be used for quick diagnosis of Schizophrenia. The model has been trained using supervised learning regime. Baseline classifier ZeroR and Naive Bayes classifier have been also built using the same dataset. The performance parameters of these classifiers have been used for external validation of cost-sensitive J48 model, and metrics such as sensitivity, specificity, receiver operating characteristic curves, and Diagnostic Odd Ratio are used for analyzing the performances. The cost-sensitive J48 model developed in this study has performed at 78% accuracy, with specificity 57.4%, sensitivity 89.7% and area under ROC 0.895. The model has low specificity thus it is not appropriate to rule in other diseases; however, its high sensitivity could reduce miss detection in Schizophrenia. Thus CS-J48 model can therefore be relied on, as an ensemble, to classify Schizophrenia diagnosis. Recommendations are suggested for future work on this study which include using larger number of records to train the model better in the future for higher accuracy, developing the model for mobile application for public use, and exploring other data mining algorithms such as deep learning neural network.

Conflict of interests: The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement: The research received no funding from any grant or donor organization.

References

1. Koch SP, Hägele C, Haynes J-D, et al. Diagnostic classification of schizophrenia patients on the basis of regional reward-related fMRI signal patterns. PLoS One. 2015;10:e0119089.
2. Han J, Kamber M, Pei J. Data mining: concepts and techniques. (3rdedn), Morgan Kaufmann, 2012;pp.6-8.
3. Hand DJ. Data Mining. (2ndedn), Encyclopedia of Environmetrics. John Wiley & Sons, Ltd., 2012;pp.1-4.
4. Jothikumar R, Sivabalan R, Sivarajan E. Accuracies f J48 WEKA classifier with different supervised WEKA filters for predicting heart diseases. ARPN J Eng Appl Sci. 2015;10:7788-93.
5. Gholap J. Performance tuning of j48 algorithm for prediction of soil fertility. Asian J Comput Sci Inf Technol. 2012;2.
6. Yadav AK, Chandel SS. Solar energy potential assessment of western Himalayan Indian state of Himachal Pradesh using J48 algorithm of WEKA in ANN based prediction model. Renew Energy. 2015;75:675-93.
7. Aljawarneh S, Aldwairi M, Yassein MB. Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model. J Comput Sci-Neth. 2018;25:152–60.

8. Kaur G, Chhabra A. Improved J48 classification algorithm for the prediction of diabetes. Int J Comput Appl. 2014;98:13-7.
9. Daraei A, Hamidah H. An efficient predictive model for myocardial infarction using cost-sensitive J48 model. Iran J Public Health. 2017;46:682–92.
10. Dvey-Aharon Z, Fogelson N, Peled A, et al. Schizophrenia detection and classification by advanced analysis of EEG recordings using a single electrode approach. PLoS One. 2015;10:e0123033
11. Iwabuchi SJ, Liddle PF, Palaniyappan L. Clinical utility of machine-learning approaches in schizophrenia: improving diagnostic confidence for translational neuroimaging. Front Psychiatry. 2013;4:95.
12. Dabek F, Caban JJ. A neural network based model for predicting psychological conditions. In: Guo Y, Friston K, Aldo F, Hill S, Peng H. (eds) Brain Informatics and Health. LNAI 9250, Springer, Cham. 2015;pp. 252–61.
13. Shim M, Hwang H J, Kim DW, et al. Machine-learning-based diagnosis of schizophrenia using combined sensor-level and source-level EEG features. Schizophr Res. 2016;176:314-9.
14. Han S, Huang W, Zhang Y, et al. Recognition of early-onset schizophrenia using deep-learning method. Appl Inform. 2017;4.
15. Devia C, Mayol-Troncoso R, Parrini J, et al. EEG classification during scene free-viewing for schizophrenia detection. IEEE Trans Neural Syst Rehabil Eng. 2019;27:1193-9.
16. https://weka.sourceforge.io/packageMetaData/
17. https://cordis.europa.eu/project/id/242114/reporting
18. Gao B, Woo WL. Wearable audio monitoring: Content-based processing methodology and implementation. IEEE T Hum-Mach Syst. 2014; 44:222-33.
19. Gu J, Gao B, Chen Y, et al. Wearable social sensing: Content-based processing methodology and implementation. IEEE Sens J. 2017;17:7167-76.
20. Chen Y, Gao B, Jiang L, et al. Transfer learning for wearable long-term social speech evaluation. IEEE Access. 2018;6:61305-16.
21. Yang S, Gao B, Jiang L, et al. IoT structured long-term wearable social sensing for mental wellbeing. IEEE Internet Things J. 2019;6:3652-62.
22. Jiang L, Gao B, Gu J, et al. Wearable long-term social sensing for mental wellbeing. IEEE Sens J. 2019; 19:8532-42.
23. Jin J, Gao B, Yang S, et al. Attention-block deep learning based features fusion in wearable social sensor for mental wellbeing evaluations. IEEE Access. 2020;8:89258-68.
24. Woo WL. Future trends in I&M: Human-machine co-creation in the rise of AI. IEEE Instru Meas Mag. 2020;23:71-3.
25. Hamad RA, Yang L, Woo WL, et al. Joint learning of temporal models to handle imbalanced data for human activity recognition. Appl Sci. 2020;10:5293.
26. Woo WL, Gao B. Sensor signal and information processing II. Sensors. 2019;20:3751.

Appendix 1: Representation of features, description, and values

No	FEATURE	DESCRIPTION	VALUES
1	Y_O_REP	Year patient reported in Hospital	Year e.g. 2017
2	AGE	Age of patient	Age e.g. 32, 23
3	SEX	Sex of patient	MALE, FEMALE
4	OCCUP_HX	Occupation History	unemployed, occupation
5	MAR_STA	Marital status	married,single, divorced, widow
6	DUR_EPIS	Episode Duration (length of time the patient has suffered the symptoms	time in months
7	P_PXY_HX	Past Psychiatric History	e.g rape, mental illness, etc
8	P_MED_HX	Past Medical History	No, disease suffered in past (eg diabetes)
9	FAM_P_HX	Family Psychiatric History	Yes, No
10	P_SOC_HX	Past Social History	Yes, No
11	P_SEX_HX	Past Sexual History	normal, experience (e.g masturbate, gonorrhea,etc)
12	FOR_HX	Forensic History	No, Yes
13	PREMOB_HX	Pre-morbid History	Normal, introvert, extrovert, melancholic
14	MSE	Mental State Examination	Kempt, unkempt, poor eye contact, restless
15	SPEECH	Speech Status	Normal, reduced volume, mute, slurred, decreased tone, irrelevant, incoherent
16	MOOD	Mood of the patient at the time of report	Euthemic, neutral, happy, relaxed, fine/ok, worried, sad, irritable
17	AFFECT	Affect of the patient at time of report	Depressed, reactive, blunt, restricted, congruent, abnormal
18	TH_FORM	Thought Form at time of report	Logical, abnormal
19	TH_STRM	Thought Stream at time of report	Reduced, normal, increased
20	TH_CONTENT	Thought content at time of report	Persecutory delusion, auditory hallucination, normal, obsession, grandiose delusion, disorder
21	TH_POSSESSIO N	Thought Possession at	Impaired,
		time of report	Normal
22	PERCEP	Perception at time of report	No, Auditory Hallucination, visual hallucination, tactile hallucination, olfactory hallucination, preoccupation
23	ORIENT	Time, Place and Position Orientation at time of report	Oriented in TPP, no
24	ATTEN	Attention status at time of report	Rousable, poor
25	CONC	Concentration status at time of report	Good, reduced, poor
26	MEM_IR	Immediate Recall status at time report	Good, fair, poor
27	MEM_ST	Short-term Memory status at time report	Good, fair, poor
28	MEM_LT	Long-Term Memory status at time report	Good, fair, poor
29	INT_GFK	Intelligence Test of General Fund of Knowledge	Good, fair, poor
30	INT_S_A_D	Intelligence Test of Similarity and Difference	Good, fair, poor
31	INT_CAL	Intelligence Test of Arithmetic	Good, fair, poor
32	INT_PROV	Intelligence Test of Proverbs	Good, fair, poor
33	JUDGMT	Intellectual Judgment status at time of report	Good, poor
34	INSIGHT	Insight status at time of report	Good, partial, poor
35	PSE	Physical State Examination status	Good, normal, pale
36	EEG	Electroencephalogram (to exclude brain tumor possibility status).	Normal, altered
37	DIAGN	Result of diagnosis	Diagnosis Result e.g. Bipolar disorder
38	CLASS	Classes of instances	SCHIZ, OTHERS

Appendix 2: Resulting decision-tree using the Cost-Sensitive Pruned J48 algorithm

```
PSE = GOOD: SCHIZ (59.14)
PSE = NORMAL
|   PREMOBD_HX = NORMAL
|   |   P_MED_HX = HYPTENSIVE: SCHIZ (0.0)
|   |   P_MED_HX = SHORT-SIGHT: SCHIZ (0.0)
|   |   P_MED_HX = NO
|   |   |   MEM_IR = GOOD: SCHIZ (22.07/3.37)
|   |   |   MEM_IR = FAIR: SCHIZ (3.33/0.1)
|   |   |   MEM_IR = POOR: OTHERS (5.0/1.2)
|   |   P_MED_HX = DIABETES: SCHIZ (5.18/0.3)
|   |   P_MED_HX = SIEZURE: OTHERS (1.94)
|   |   P_MED_HX = MALARIA: SCHIZ (0.0)
|   |   P_MED_HX = HADE: SCHIZ (0.0)
|   |   P_MED_HX = SHADE
|   |   |   CONC = GOOD: OTHERS (6.76/0.75)
|   |   |   CONC = POOR: SCHIZ (4.25/0.13)
|   |   |   CONC = REDUCED: OTHERS (0.0)
|   |   P_MED_HX = BLOOD TRANS: SCHIZ (2.59/0.15)
|   |   P_MED_HX = ALLEGIES: SCHIZ (0.0)
|   |   P_MED_HX = HbSS: OTHERS (3.25/1.23)
|   |   P_MED_HX = STROKE: SCHIZ (0.0)
|   |   P_MED_HX = HDASP: OTHERS (1.94)
|   |   P_MED_HX = IMMUNE DISORDER: SCHIZ (0.0)
|   |   P_MED_HX = ASPHYXIA: OTHERS (1.94)
|   |   P_MED_HX = JAUNDICE: OTHERS (1.94)
|   |   P_MED_HX = GLAUCOMA: SCHIZ (0.0)
|   |   P_MED_HX = ASTHMA: OTHERS (1.94)
|   PREMOBD_HX = INTROVERT: SCHIZ (9.74)
|   PREMOBD_HX = MOODY: SCHIZ (0.0)
|   PREMOBD_HX = ASTHMA: SCHIZ (2.44)
|   PREMOBD_HX = MELANCHOLIC: OTHERS (1.83)
PSE = PALE
|   INT_GFK = FAIR: SCHIZ (4.49/0.52)
|   INT_GFK = GOOD: OTHERS (6.74/2.31)
|   INT_GFK = POOR: SCHIZ (4.49/0.52)


Number of Leaves: 30
Size of the tree: 36
```

Appendix 3: Resulting decision-tree using the J48 algorithm

```
PSE = GOOD: SCHIZ (48.28)
    PSE = NORMAL
    |   PREMOBD_HX = NORMAL
    |   |   TH_FORM = LOGICAL
    |   |   |   P_MED_HX = HYPTENSIVE: OTHERS (0.0)
    |   |   |   P_MED_HX = SHORT-SIGHT: OTHERS (0.0)
    |   |   |   P_MED_HX = NO
    |   |   |   |   CONC = GOOD
    |   |   |   |   |   TH_POSS = IMPAIRED: OTHERS (3.13)
    |   |   |   |   |   TH_POSS = NORMAL: SCHIZ (13.72/1.72)
    |   |   |   |   CONC = POOR: OTHERS (6.26)
    |   |   |   |   CONC = REDUCED: SCHIZ (0.0)
    |   |   |   P_MED_HX = DIABETES: SCHIZ (2.2/0.2)
    |   |   |   P_MED_HX = SIEZURE: OTHERS (3.3)
    |   |   |   P_MED_HX = MALARIA: OTHERS (0.0)
    |   |   |   P_MED_HX = HADE: OTHERS (0.0)
    |   |   |   P_MED_HX = SHADE
    |   |   |   |   CONC = GOOD: OTHERS (11.04/0.86)
    |   |   |   |   CONC = POOR: SCHIZ (3.41/0.14)
    |   |   |   |   CONC = REDUCED: OTHERS (0.0)
    |   |   |   P_MED_HX = BLOOD TRANS: SCHIZ (1.84/0.17)
    |   |   |   P_MED_HX = ALLEGIES: OTHERS (0.0)
    |   |   |   P_MED_HX = HbSS: OTHERS (3.3)
    |   |   |   P_MED_HX = STROKE: OTHERS (0.0)
    |   |   |   P_MED_HX = HDASP: OTHERS (3.3)
    |   |   |   P_MED_HX = IMMUNE DISORDER: OTHERS (0.0)
    |   |   |   P_MED_HX = ASPHYXIA: OTHERS (2.76)
    |   |   |   P_MED_HX = JAUNDICE: OTHERS (2.76)
    |   |   |   P_MED_HX = GLAUCOMA: OTHERS (0.0)
    |   |   |   P_MED_HX = ASTHMA: OTHERS (3.3)
    |   |   TH_FORM = ABNORMAL: SCHIZ (10.77/1.34)
    |   |   TH_FORM = NFTD: SCHIZ (1.18/0.15)
    |   PREMOBD_HX = INTROVERT: SCHIZ (8.0)
    |   PREMOBD_HX = MOODY: OTHERS (0.0)
    |   PREMOBD_HX = ASTHMA: SCHIZ (2.0)
    |   PREMOBD_HX = MELANCHOLIC: OTHERS (3.0)
    PSE = PALE
    |   INT_GFK = FAIR: SCHIZ (3.49/0.6)
    |   INT_GFK = GOOD: OTHERS (10.48/2.68)
    |   INT_GFK = POOR: SCHIZ (3.49/0.6)


    Number of Leaves:       33
    Size of the tree:       41
```

Appendix 4: Naive Bayes Model

Attribute	Class SCHIZ (0.64)	OTHERS (0.36)
OCCUP		
NURSE	4.0	1.0
UNEMPLOYED	29.0	10.0
STUDENT	27.0	22.0
CLEANER	4.0	1.0
SEA MISTRESS	4.0	1.0
TEACHING	4.0	1.0
TEACHER	3.0	1.0
SALESWOMAN	3.0	1.0
TRADER	9.0	10.0
GRADUATE	3.0	1.0
PHYSIOTHERAPIST	3.0	1.0
ARCHITECT	3.0	1.0
GUARD	3.0	1.0
ASSISTANT	3.0	1.0
RETIRED	4.0	1.0
ENGINEER	1.0	4.0
WRITER	1.0	4.0
COBBLER	1.0	4.0
AUDITOR	1.0	4.0
LAWYER	1.0	4.0
[total]	111.0	74.0
MAR_STA		
WIDOW	9.0	1.0
MARRIED	24.0	16.0
SINGLE	62.0	37.0
DIVORCED	4.0	4.0
[total]	99.0	58.0
DUR_EPIS		
mean	51.2503	75.5045
std. dev.	80.8657	115.2294
weight sum	95	53
precision	16.6739	16.6739
P_PSY_HX		
RAPE	4.0	1.0
NO	35.0	10.0
SUSPISION	4.0	1.0
MENTAL ILLNESS	42.0	40.0
IRRITABILITY	3.0	1.0
GRIEF	1.0	4.0
[total]	89.0	57.0
P_MED_HX		
HYPTENSIVE	4.0	1.0
SHORT-SIGHT	4.0	1.0
NO	35.0	13.0
DIABETES	11.0	1.0
SIEZURE	3.0	4.0
MALARIA	3.0	1.0

HADE	3.0	1.0
SHADE	6.0	10.0
BLOOD TRANS	3.0	1.0
ALLEGIES	3.0	1.0
HbSS	5.0	4.0
STROKE	3.0	4.0
HDASP	3.0	4.0
IMMUNE DISORDER	1.0	4.0
ASPHYXIA	1.0	4.0
JAUNDICE	1.0	4.0
GLAUCOMA	1.0	4.0
ASTHMA	1.0	4.0
[total]	91.0	66.0
FAM_P_HX		
NO	61.0	36.0
YES	20.0	17.0
[total]	81.0	53.0
P_SOC_HX		
YES	34.0	18.0
NO	52.0	38.0
[total]	86.0	56.0
P_SEX_HX		
NORMAL	73.0	46.0
MASTURBATE	1.0	4.0
GORNORREA	1.0	4.0
[total]	75.0	54.0
FOR_HX		
NO	72.0	51.0
YES	10.0	2.0
[total]	82.0	53.0
PREMOBD_HX		
NORMAL	54.0	49.0
INTROVERT	16.0	1.0
MOODY	3.0	1.0
ASTHMA	3.0	1.0
MELANCHOLIC	1.0	4.0
[total]	77.0	56.0
MSE		
KEMPT	75.0	42.0
POOR EYE CONTACT	13.0	7.0
UNKEMPT	10.0	5.0
RESTLESS	3.0	4.0
[total]	101.0	58.0
SPEECH		
NORMAL	62.0	38.0
REDUCED VOL	4.0	1.0
IRRELEVANT	9.0	1.0
INCOHERENT	3.0	1.0
DECREASED TONE	15.0	7.0
MUTE	3.0	6.0
SLURRED	1.0	4.0
INCREASED TONE	1.0	4.0
[total]	98.0	62.0
MOOD		

NEUTRAL	8.0	7.0
HAPPY	32.0	19.0
RELAXED	4.0	1.0
FINE	14.0	7.0
OK	21.0	7.0
WORRIED	5.0	1.0
SAD	9.0	10.0
EUTHYMIC	5.0	7.0
IRRITABLE	3.0	1.0
[total]	101.0	60.0
AFFECT		
DEPRESSED	6.0	7.0
REACTIVE	35.0	22.0
BLUNT	23.0	4.0
RESTRICTED	18.0	4.0
ABNORMAL	4.0	2.0
IRRITABLE	5.0	4.0
FEARFUL	3.0	1.0
CONGRUENT	7.0	10.0
SUSPICIOUS	3.0	1.0
[total]	104.0	55.0
TH_FORM		
LOGICAL	46.0	43.0
ABNORMAL	21.0	1.0
NFTD	2.0	1.0
[total]	69.0	45.0
TH_STRM		
REDUCED	18.0	1.0
NO	4.0	1.0
NORMAL	45.0	37.0
INCREASED	7.0	7.0
[total]	74.0	46.0
TH_CONTENT		
PERSECUTORY DELUSION	40.0	13.0
NORMAL	25.0	22.0
OBSESSION	6.0	4.0
AUDITORY HALLUCINATION	4.0	1.0
DELUSION	6.0	1.0
DELUSION OF REFERENCE	3.0	1.0
GRANDIOSE DELUSION	3.0	1.0
AUDI TORY DELUSION	3.0	1.0
DISORDER	1.0	4.0
GRANDEUR DELUSION	1.0	4.0
[total]	92.0	52.0
TH_POSS		
IMPAIRED	8.0	7.0
NORMAL	45.0	37.0
[total]	53.0	44.0
PERCEP		
AUDITORY HALLUCINATION	35.0	7.0
FUNCTIONAL HALLUCINATION	4.0	1.0
SOMATIC HALLUCINATION	6.0	1.0
TACTILE HALLUCINATION	5.0	4.0
OLFACTOTY HALLUCINATION	3.0	1.0
HALLUCINATION	7.0	1.0
VISUAL HALLUCINATION	7.0	4.0
PREOCCUPATION	3.0	1.0
NO	12.0	28.0
[total]	82.0	48.0
ORIENT		
ORIENTED IN TPP	77.0	43.0
IMPAIRED	3.0	1.0

[total]	80.0	44.0
CONC		
GOOD	63.0	37.0
POOR	16.0	7.0
REDUCED	8.0	1.0
[total]	87.0	45.0
MEM_IR		
GOOD	68.0	37.0
FAIR	6.0	1.0
POOR	3.0	7.0
[total]	77.0	45.0
INT_GFK		
FAIR	12.0	1.0
GOOD	34.0	34.0
POOR	7.0	7.0
[total]	53.0	42.0
JUDGMT		
POOR	42.0	25.0
GOOD	35.0	22.0
[total]	77.0	47.0
INSIGHT		
PARTIAL	27.0	22.0
POOR	41.0	13.0
PERSISTENT	4.0	1.0
GOOD	18.0	13.0
[total]	90.0	49.0
PSE		
GOOD	48.0	1.0
NORMAL	39.0	46.0
PALE	9.0	10.0
[total]	96.0	57.0
EEG		
NORMAL	98.0	49.0
ALTERED	1.0	7.0
[total]	99.0	56.0