LOVELY PROFESSIONAL UNIVERSITY

NAME: KANISH SAINI

MCA HONS.(DATA SCIENCE)

REGNO. 12321902

SECTION : D2340

SUBMITTED TO : MS. AMNINDER KAUR

TOPIC

Computer-aided decision-making for predicting liver disease using rapid miner

TABLE OF CONTENT

Abstract

Using medical data mining models has been considered as a significant way to predict diseases in recent years.

In the field of healthcare, we face a large amount of data, and this is one of the challenges in predicting and analyzing the target disease.

With the help of data mining models, one can convert this data into valuable information, and through analyzing them logically and scientifically, one can reach accurate decision-making and actual prediction.

Another challenge in the field of disease prediction is selecting features that are more significant than other features.

Feature subset selection is performed to improve the performance of models with the highest accuracy.

The purpose of this study is to select significant features by comparing data mining models to predict liver disease based on an extraction, loading, transformation, analysis (ELTA) approach for correct diagnosis.

Hence, the data mining models are compared based on the ELTA approach, such as random forest, Multi-Layer Perceptron (MLP) neural network, Bayesian networks, Support Vector Machine (SVM), and Particle Swarm Optimization (PSO)-SVM.

Among these models, the PSO-SVM model has the best performance regarding the criteria of specificity, sensitivity, accuracy, Area under the Curve (AUC), F-measure, precision, and False Positive Rate (FPR).

Furthermore, a 10-fold cross-validation method for evaluation of models is used so that the models were evaluated on a liver disease dataset.

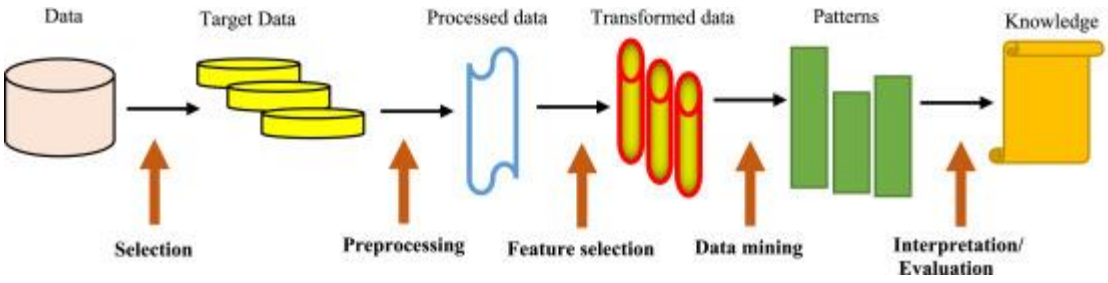
The average of estimated accuracy was calculated as 87.35%, 78.91%, 66.78%, 76.51% and 95.17% for Random forest, MLP Neural network, Bayesian network, SVM and PSO-SVM models, respectively.

Regarding the mentioned evaluation criteria, we obtained the highest performance of accuracy with the least number of features through the model.

Introduction

In medicine, improving the quality of healthcare can be better defined by the driving forces that affect it, including healthcare data; in other words, in any patient-centered quality improvement program, the data are counted as the center of that program . Extracting knowledge from the data associated with medical records by using the data mining process can lead to the identification of the laws governing the creation, growth, and acceleration of illnesses, and provide valuable information to identify the causes of disease diagnosis and treatment according to the environmental factors prevailing in healthcare.

Data mining methods can be used for identifying disease in discovering the knowledge process. Hence, data mining can discover hidden relationships, trends, and patterns among data, leading to the enhancement of the accurate identification of the disease . The concept of Knowledge discovery in databases (KDD) is a basis including theories and methods used to extract knowledge from data. KDD includes a five step process (selection, preprocessing, transformation, data mining, and interpretation/evaluation) as demonstrated in.



Since the case study of this paper is a diagnosis of liver disease, we explain as follows. The liver is the second most significant internal organ in the human body, playing a significant role in metabolism and serving several vital functions, e.g. decomposition of red blood cells .

Usually, more than 75% or three-quarters of liver tissue needs to be affected before a decrease in function occurs. Between 1999 and 2016, fatalities from cirrhosis incremented by 65% to 34174 in the United States, while deaths from hepatocellular carcinoma doubled to more than 11073, according to findings published in the British Medical Journal .

Only a segment of Asia-Pacific Islanders observed a recuperation in the annual death rate from hepatocellular carcinoma per year. An increment in the fatality rate from cirrhosis was more significant for Native Americans. Young people (25–34 years old) experienced the highest average annual increase in cirrhosis-related deaths, which was driven by alcohol-related liver disease.

The United States in the South and West exhibited a sudden increase in annual deaths from cirrhosis. There are specific factors for the emergence of liver disease, the most significant of which are a family history of liver disease, smoking, alcohol consumption, obesity, and diabetes . Predictive and descriptive models are applied to medical data mining as data mining techniques.

Medical imaging is used to diagnose liver disease, such a Sonography, CT scan, and MRI. These tools are associated with some harmful effects and high costs. Hence, researchers have proposed methods to replace imaging devices in the diagnosis of disease, the most discussed of which is data mining.

The purpose of this study is based on the proposed data mining process called extraction, loading, transformation and analysis (ELTA) to compare multiple prediction models for liver disease incidence.

In this study, five widely-used data mining classification models: Random Forest, MLP-Neural Network, Bayesian network, SVM and PSO-SVM, along with a 10-fold cross-validation method, were compared. Accuracy, sensitivity, specificity, (ROC, receiver operating characteristic) namely (AUC, the area under the curve), F-measure, FPR, and precision were used to evaluate them. Since the primary purpose of this study is to select the significant symptoms of liver disease.

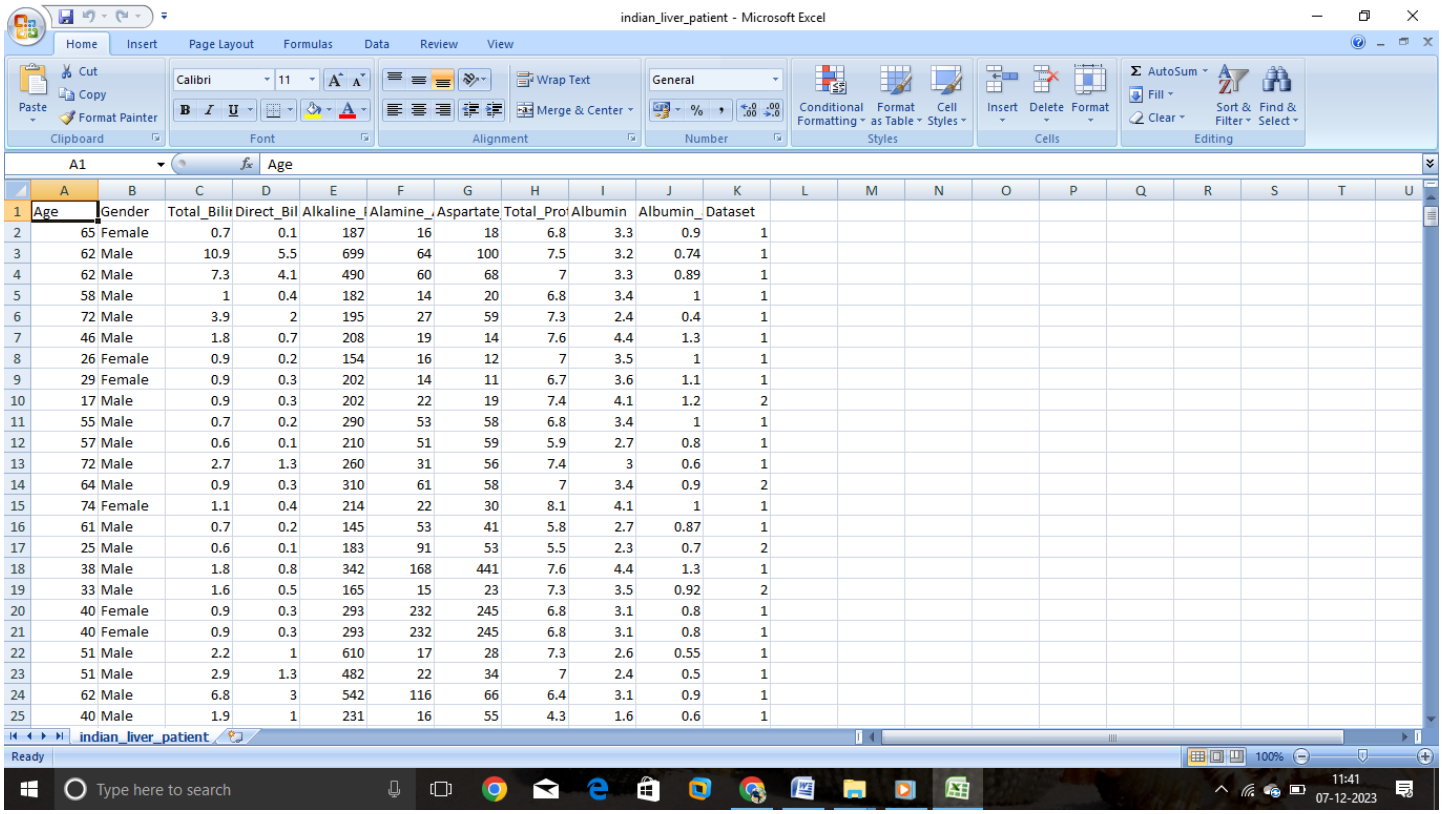
The remaining portions of the paper are organized as follows: The present study expresses the required background information , Related works discussed , The proposed methodology is presented , the performance of the models is evaluated and analyzed.

The results of the experiment. Finally Results and discussion, Conclusion and future works present findings of the research and the conclusions, namely “Results and Discussion” and “Conclusion and Future works”, respectively.

Data set :

The data set we have used here is Indian_liver_patient.csv

With 11 rows and 584 columns



Age	Gender	Total_Bilir	Direct_Bil	Alkaline_I	Alamine_A	Aspartate	Total_Pro	Albumin	Albumin_	Dataset
65	Female	0.7	0.1	187	16	18	6.8	3.3	0.9	1
62	Male	10.9	5.5	699	64	100	7.5	3.2	0.74	1
62	Male	7.3	4.1	490	60	68	7	3.3	0.89	1
58	Male	1	0.4	182	14	20	6.8	3.4	1	1
72	Male	3.9	2	195	27	59	7.3	2.4	0.4	1
46	Male	1.8	0.7	208	19	14	7.6	4.4	1.3	1
26	Female	0.9	0.2	154	16	12	7	3.5	1	1
29	Female	0.9	0.3	202	14	11	6.7	3.6	1.1	1
17	Male	0.9	0.3	202	22	19	7.4	4.1	1.2	2
55	Male	0.7	0.2	290	53	58	6.8	3.4	1	1
57	Male	0.6	0.1	210	51	59	5.9	2.7	0.8	1
72	Male	2.7	1.3	260	31	56	7.4	3	0.6	1
64	Male	0.9	0.3	310	61	58	7	3.4	0.9	2
74	Female	1.1	0.4	214	22	30	8.1	4.1	1	1
61	Male	0.7	0.2	145	53	41	5.8	2.7	0.87	1
25	Male	0.6	0.1	183	91	53	5.5	2.3	0.7	2
38	Male	1.8	0.8	342	168	441	7.6	4.4	1.3	1
33	Male	1.6	0.5	165	15	23	7.3	3.5	0.92	2
40	Female	0.9	0.3	293	232	245	6.8	3.1	0.8	1
40	Female	0.9	0.3	293	232	245	6.8	3.1	0.8	1
51	Male	2.2	1	610	17	28	7.3	2.6	0.55	1
51	Male	2.9	1.3	482	22	34	7	2.4	0.5	1
62	Male	6.8	3	542	116	66	6.4	3.1	0.9	1
40	Male	1.9	1	231	16	55	4.3	1.6	0.6	1

Literature Review:

P. Kuppan et al. [10] in their research work authors have worked on doing an analysis of the data related to Liver Disorder with the help of Naive Bayes, Decision Table, and J48. However, attributes like case history of the patient, diabetes, smoking, obesity, alcohol intake, smoking etc were used. Based upon the given database it has concluded Jagdeep Singh et al. / Procedia Computer Science 167 (2020) 1970–1980 1973 4 Jagdeep Singh et al. / Procedia Computer Science 00 (2019) 000–000 that male people are having more liver disorder than the females. Age group of 35-65 is mostly affected and out of these 26% people are having the disorder because of alcohol, smoking contributed to 22% of people, obesity, and diabetic of 4 & 5 percent respectively.

A. Gulia et al. [11] in their proposed work researchers have done classification of the liver patient data using the algorithms like Bayesian Network, Support Vector Machine, J48, Multi-Layer Perceptron and Random Forest.

The data from the UCI repository which is afforded by Center of Machine Learning and Intelligent Systems has used.

After completion of their three-phase analysis, the Random Forest Algorithm is the best one with an accuracy of 71.87% has been concluded. Y. Kumar et al. [12] in their proposed work researchers have used Rule-Based Classification Model (RBCM) for the prediction of liver diseases.

Without the rule-based classification the efficiency of all the common algorithm decreases was analyzed. In their proposed work 20 rules were used for the classification of liver diseases.

The decision tree-based algorithm gives the best performance using rule-based classification and accordingly its accuracy decreases when rule-based is not used. M. Pasha et al. [13] work on the dataset from the UCI repository is used which is having 583 instances, the metalearning algorithms like Grading, logit boost, Adaboost, and Bagging were used.

The comparisons of the algorithms based upon the amount of correct and incorrect classifications and time of execution have done.

After doing detailed analysis the grading is the best algorithm in terms of accuracy and execution time have been concluded. M. Abdar et al. [14] in their research work focuses on the early prediction of liver disease using Multilayer Perceptron Algorithm (MLPNN) which uses (CART) (classification and regression tree, (CHAID) Chi-square Automatic interaction detector, See5(C5.0). Their dataset is from UCI repository of the University of California, Irvine relevant to Indian Liver Patient Dataset (ILPD). From their results, it can be concluded that MLPNNBCHAID is the best algorithm with an innovative accuracy of 14.57%. The 70% of the data as a training data and rest of the 30% for the testing stage were used. Author EI-Shafeiy at al. [15] in their research work focuses on electronic health records, metabolomics analyses are some of the digital information related to the health and with the passage of time, the shape of Big Data has been taken. In the present work dataset having 23 attributes of 7000 patients with 5295 as male and rest as female is used.

Their proposed work make use of Boosted C5.0, Support Vector Machine, Naïve Bayes (NB) along with the feature selection.

Vijayarani et al. [16] in their research paper classification algorithms are used for the prediction of liver diseases. Famous algorithms like Naïve Bayes and Support Vector Machine (SVM) are used in the proposed work.

Hartatik, Mohammad Badri Tamam, Arief Setyanto, have examined to conclude; based on the findings of utilising the python application to test the Naive Bayes and KNN algorithms to solve predicting issues for patients with liver illness The Indian Liver Patient Dataset was obtained from the UCI Machine Learning Repository (ILPD).

The results reveal that by employing six variables in the prediction model, the Naive Bayes algorithm produces a better value than the KNN, resulting in an increase in accuracy when compared to the results of earlier studies.

Different classification techniques, such as Logistic Regression, Support Vector Machine, and K-Nearest Neighbor, were utilised by Thirunavukkarasu K,Ajay S. Singh,Md Irfan,Abhishek Chowdhur in their study to predict liver illness.

All of these algorithms were compared based on classification accuracy, which was determined using a confusion matrix. Logistic Regression and K-Nearest Neighbour have the highest accuracy, but logistic regression has the highest sensitivity, according to the experiment. As a result, we can conclude that Logistic Regression is a good way to predict liver illness

PROPOSED METHODOLOGY:

The description of each part is as follows:

A. Data Collection Selection of data is essential for selecting significant records for the analysis and to obtain productive or constructive knowledge by performing various data mining technique

B. Data Exploration Data exploration is an early step in data analysis that is used to summarise data for analysis and then to observe initial patterns of data and features. To identify highly linearly dependent features, several display approaches such as

histogram and boxplot are utilised to the extreme and outlier values feature correlation values.

C. Data Preprocessing Imputation of Missing Values –
This technique is used

- for obtaining the missing values from the data and imputating the null values with the median. In the Indian liver disease patients dataset, there are four missing values for Albumin and Globulin ratio that has been restored by median values [11].

Dummy Encoding –

Dummy encoding is a method of transforming the categorical variable to numerical variable as most of the machine algorithms are designed to work on numerical data. For each of the categorical variable, k-1 numerical variables are created. Elimination of Duplicate Values –

It is necessary to

- discard the redundant values from the data, in order to improve the efficiency and quality of the data [12]. Outlier Detection and Elimination- Outliers are

- exceptional values that differ from the remainder of the results due to minor measurement or experimental error. Outliers are divided into two categories: univariate and multivariate outliers. Whereas a single feature is considered in a univariate outlier, a multivariate outlier considers n-dimensions of ILPD data features or attributes. Resampling- The linear dataset is unbalanced, with

- the majority of patients suffering from liver illness and a small number of non-linear individuals. SMOTE is used to balance the data by synthesising additional samples for the minority class [13].

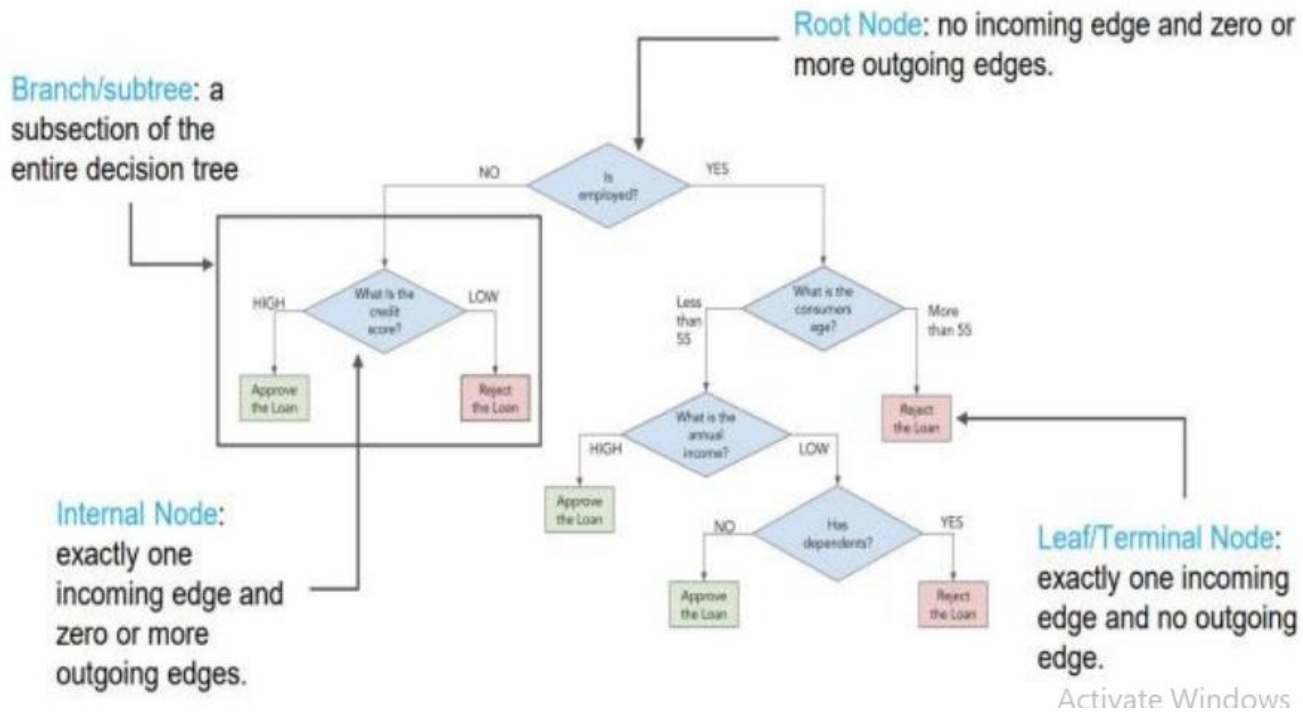
D. Feature Selection The process of limiting the number of input variables in order for the machine learning algorithm to train the model more faster is known as feature selection.

It reduces the computational complexity and makes it easier to interpret. Random Forest feature selection – Feature selection using Random forest provides highly accurate, low overfitting and easy interpretability by deriving the importance of each feature on the decision tree.

Random forest chooses features at random from decision trees created by extracting observations from the dataset at random [14].

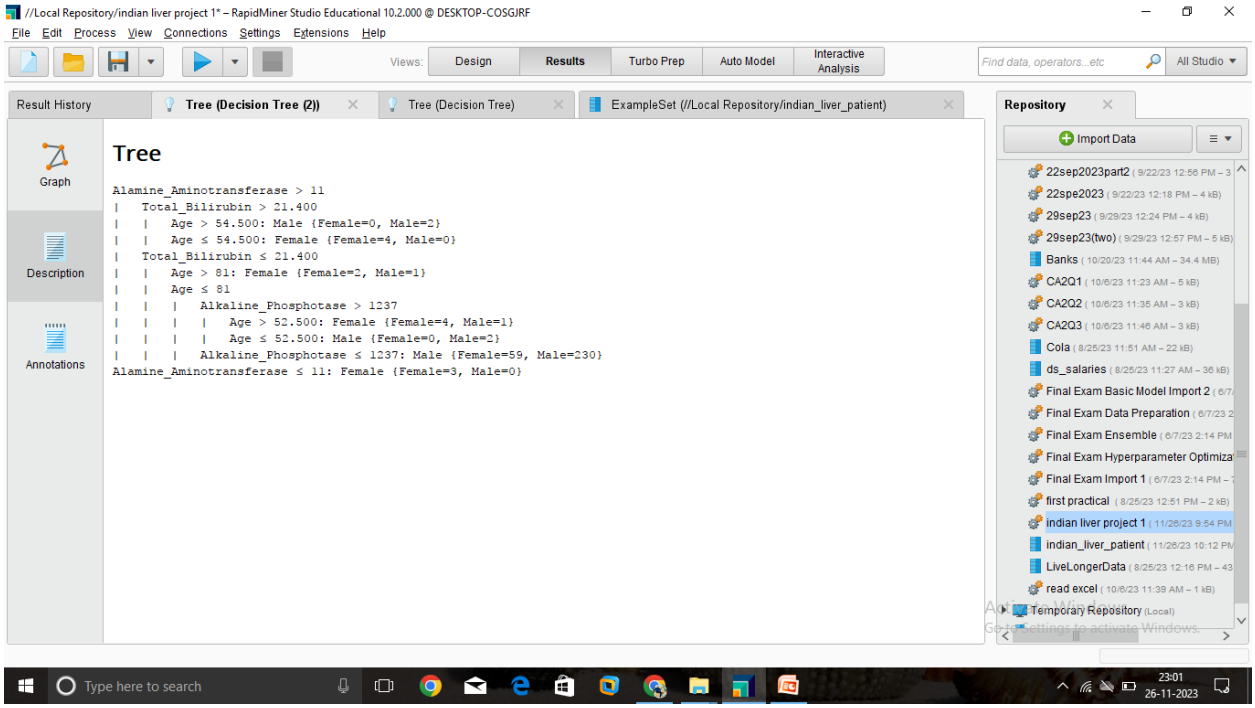
This ensures that the trees are de-correlated, making over-fitting less likely. Each tree represents a condition based on a single or more attributes.

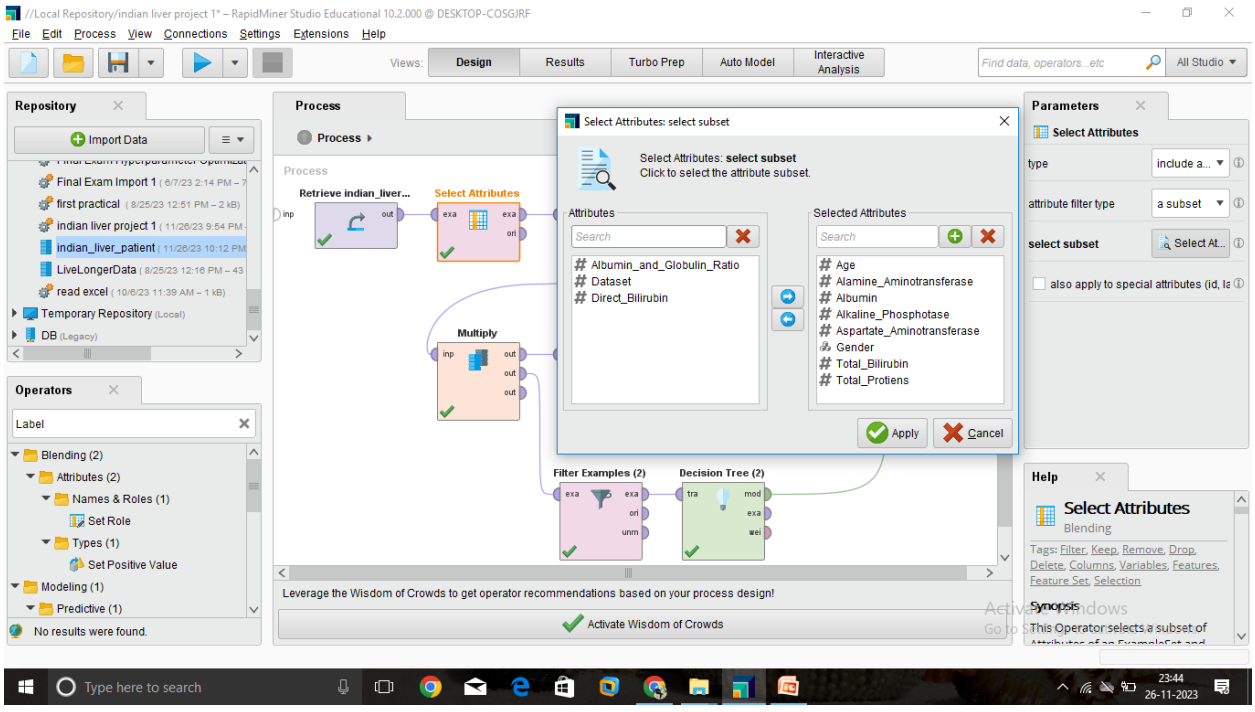
The tree separates into two buckets at each node, each of which contains observations that are more similar to one another and distinct from those in the other bucket. As a result, the value of a feature is determined by how "pure" or "impure" each bucket is.



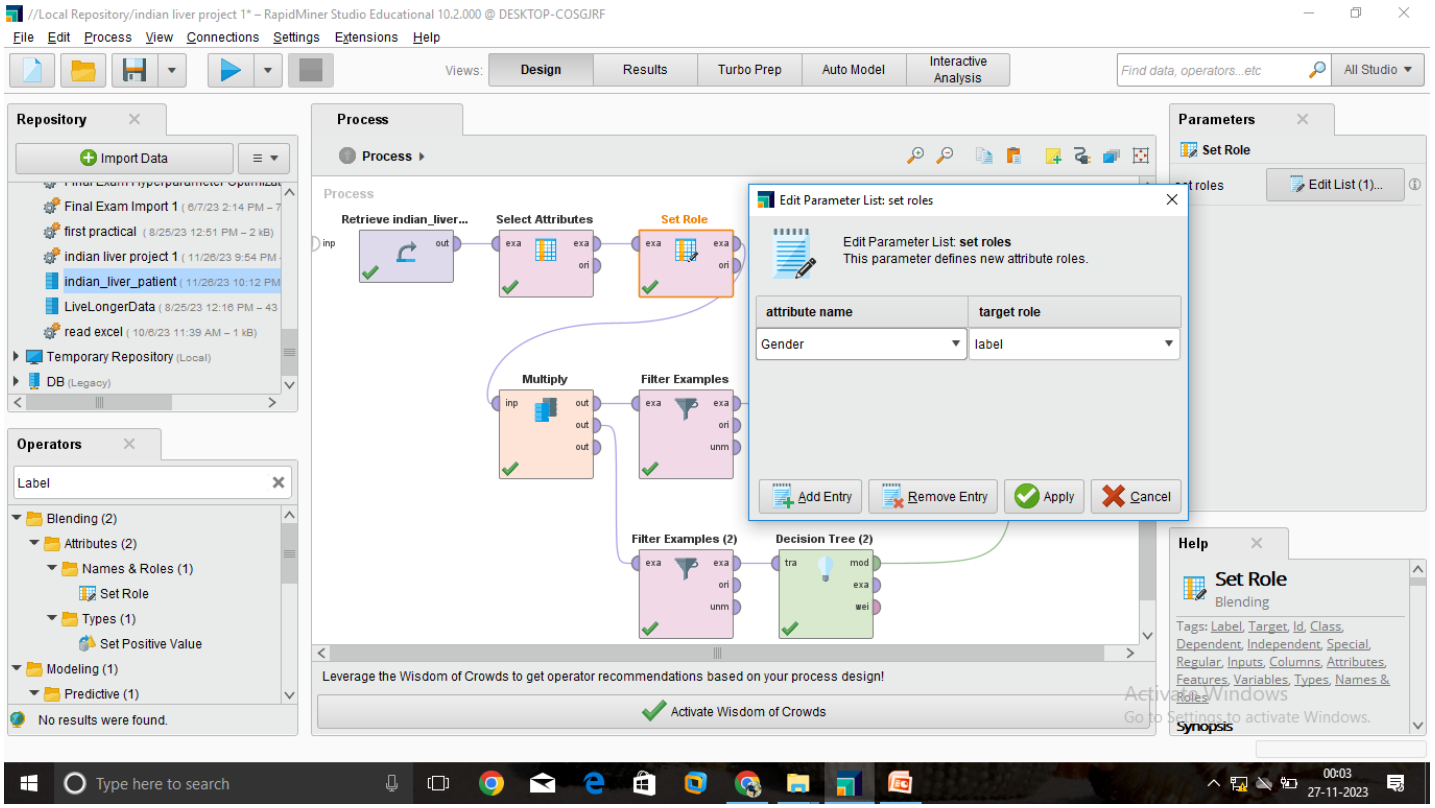
Basic Model:

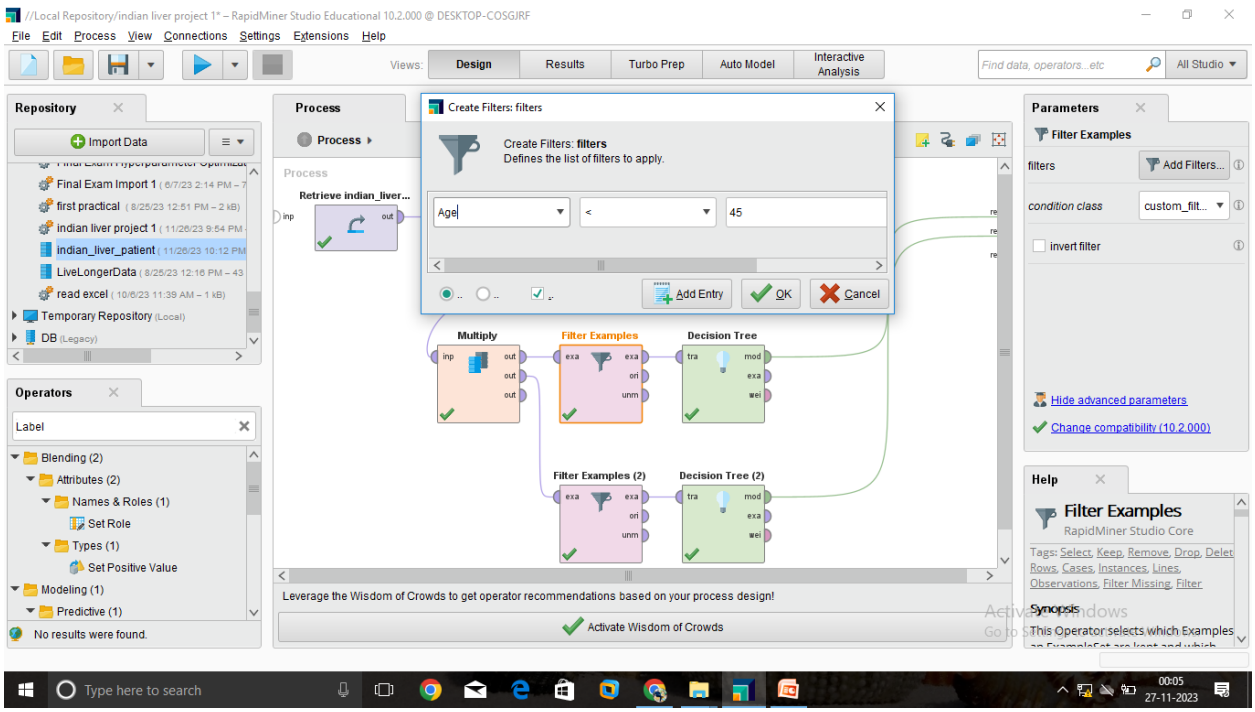
LIVER DISEASE PRIDICTION IN DIFFERENT AGE GROUPS



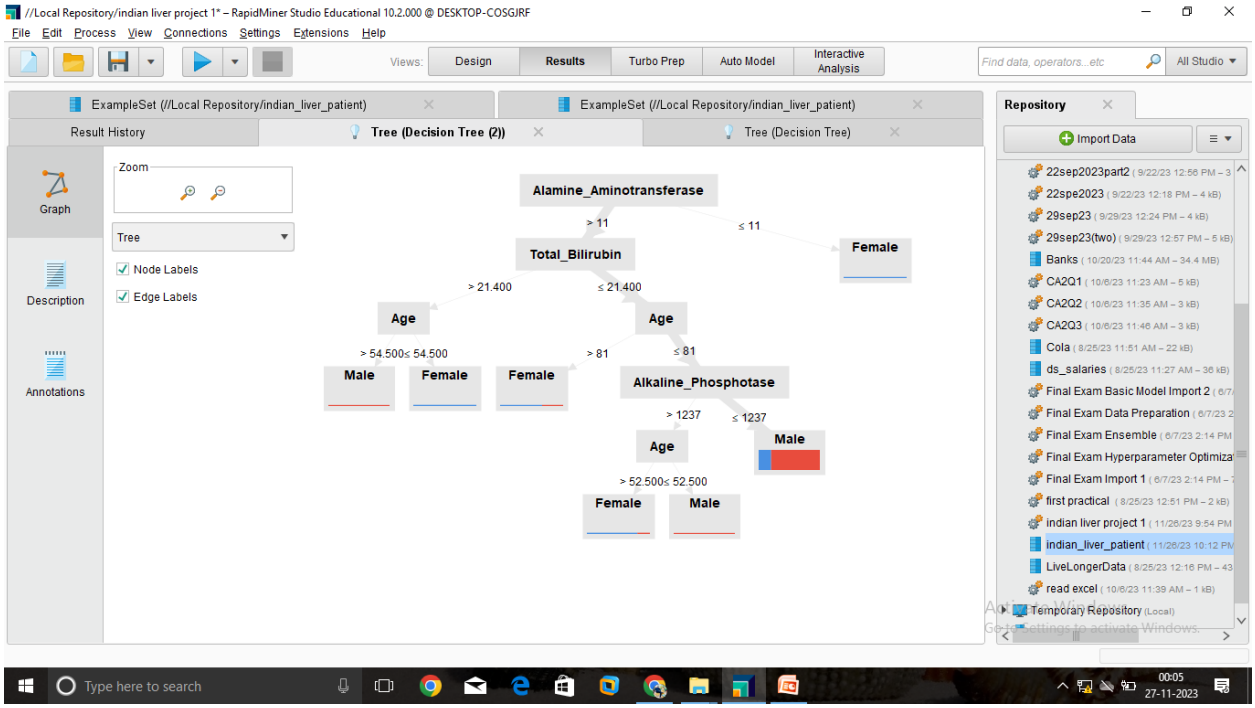


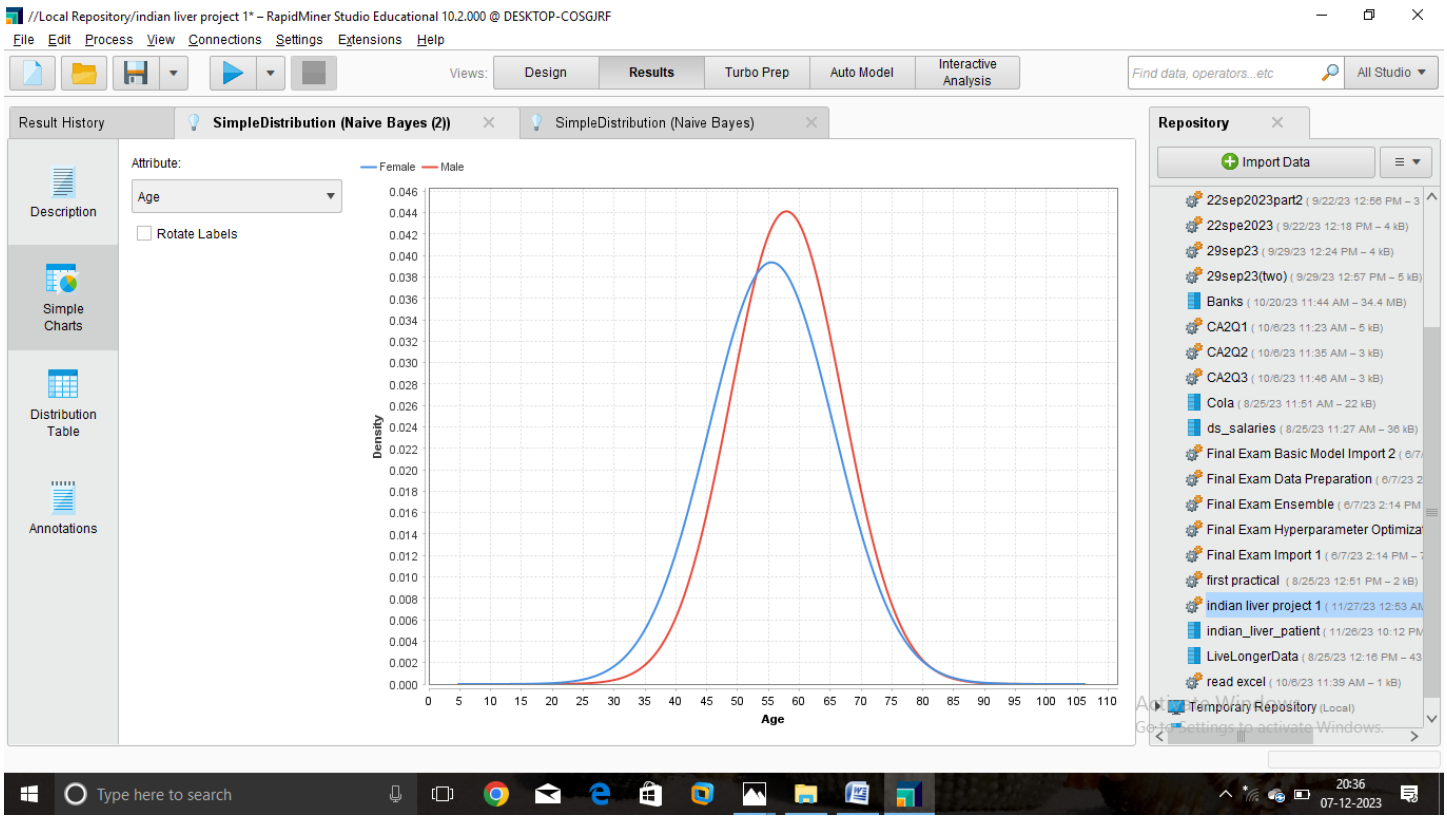
LIVER DISEASE PRIDICTION IN DIFFERENT AGE GROUPS



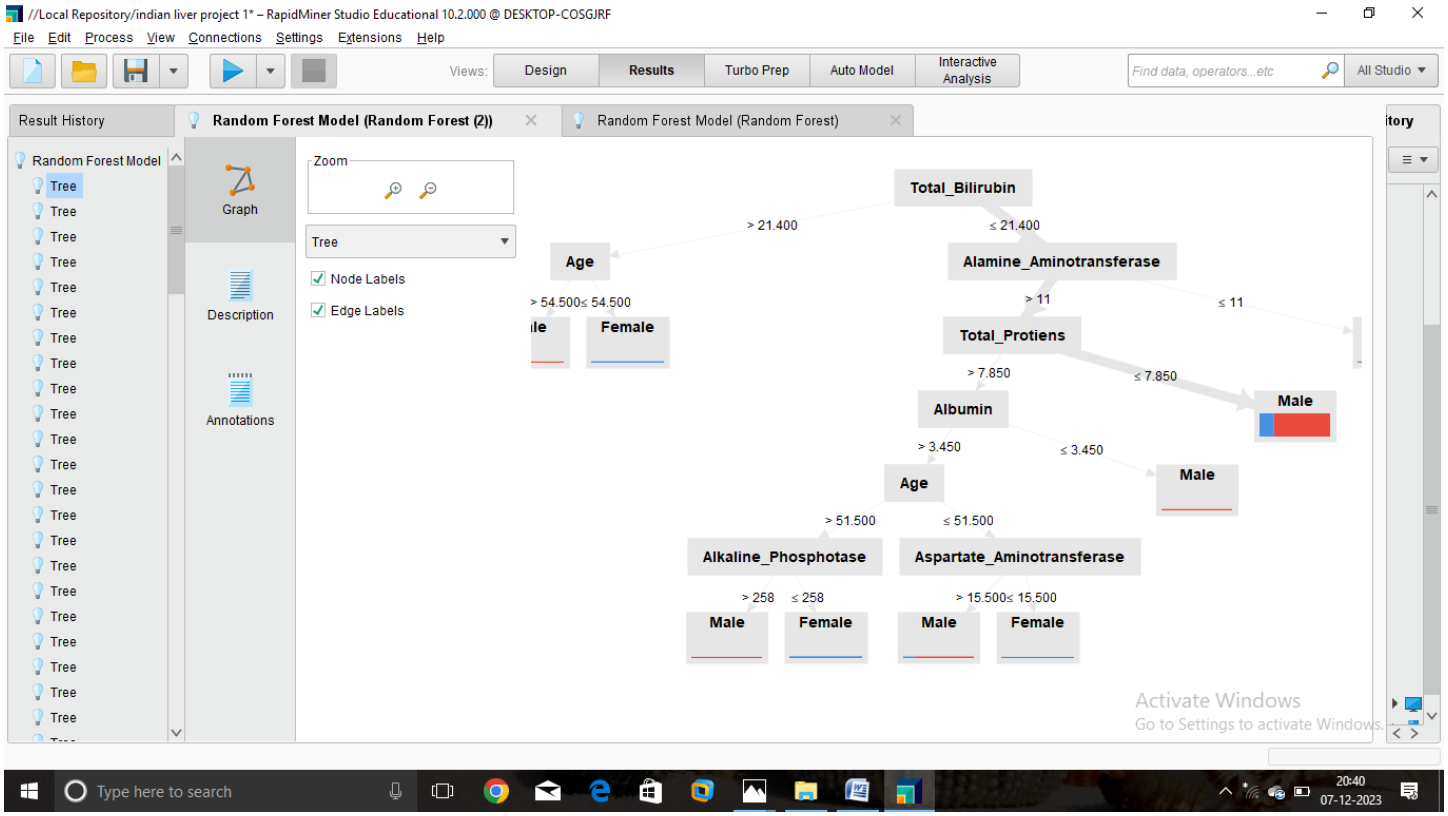


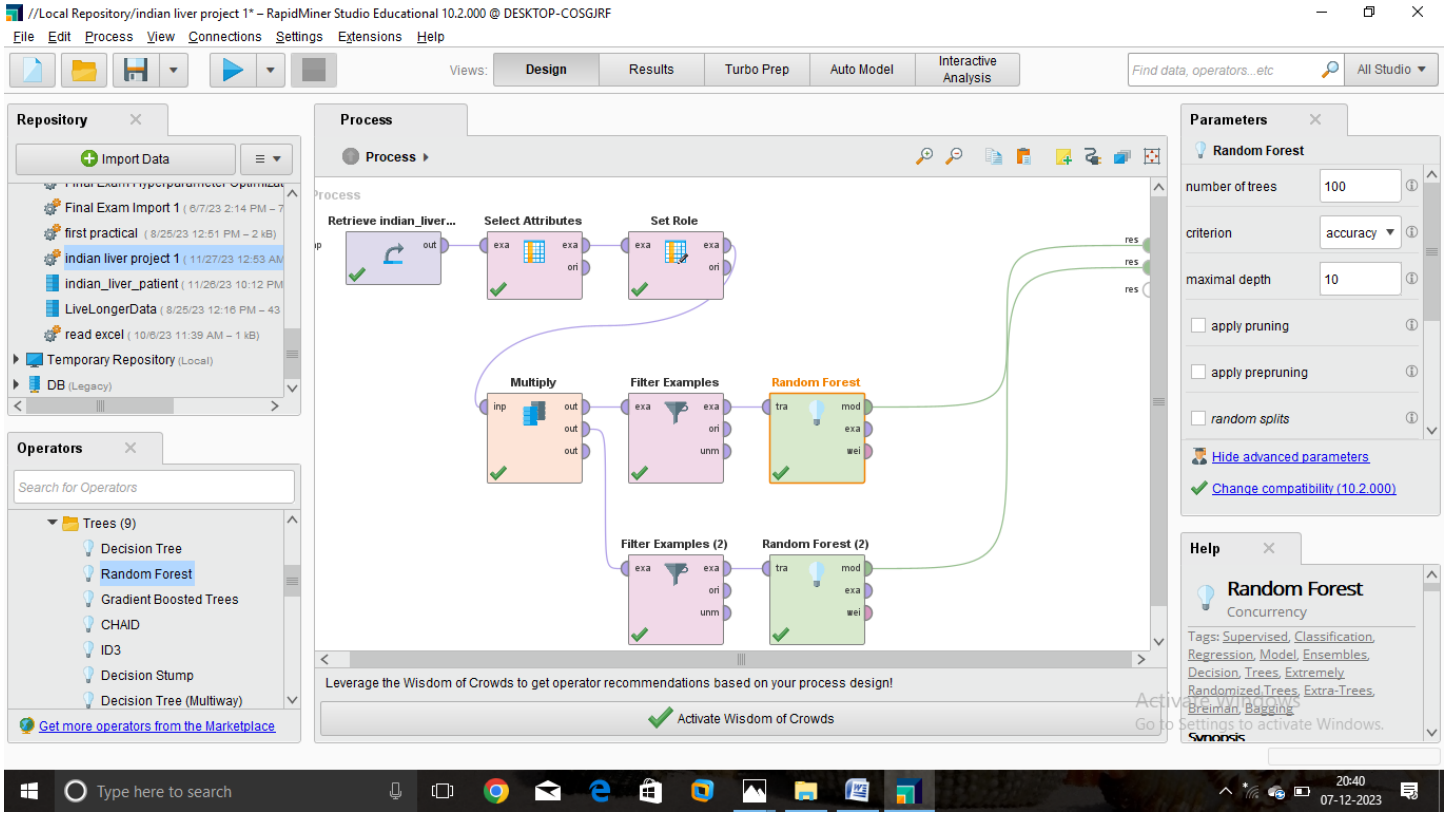
LIVER DISEASE PRIDICTION IN DIFFERENT AGE GROUPS



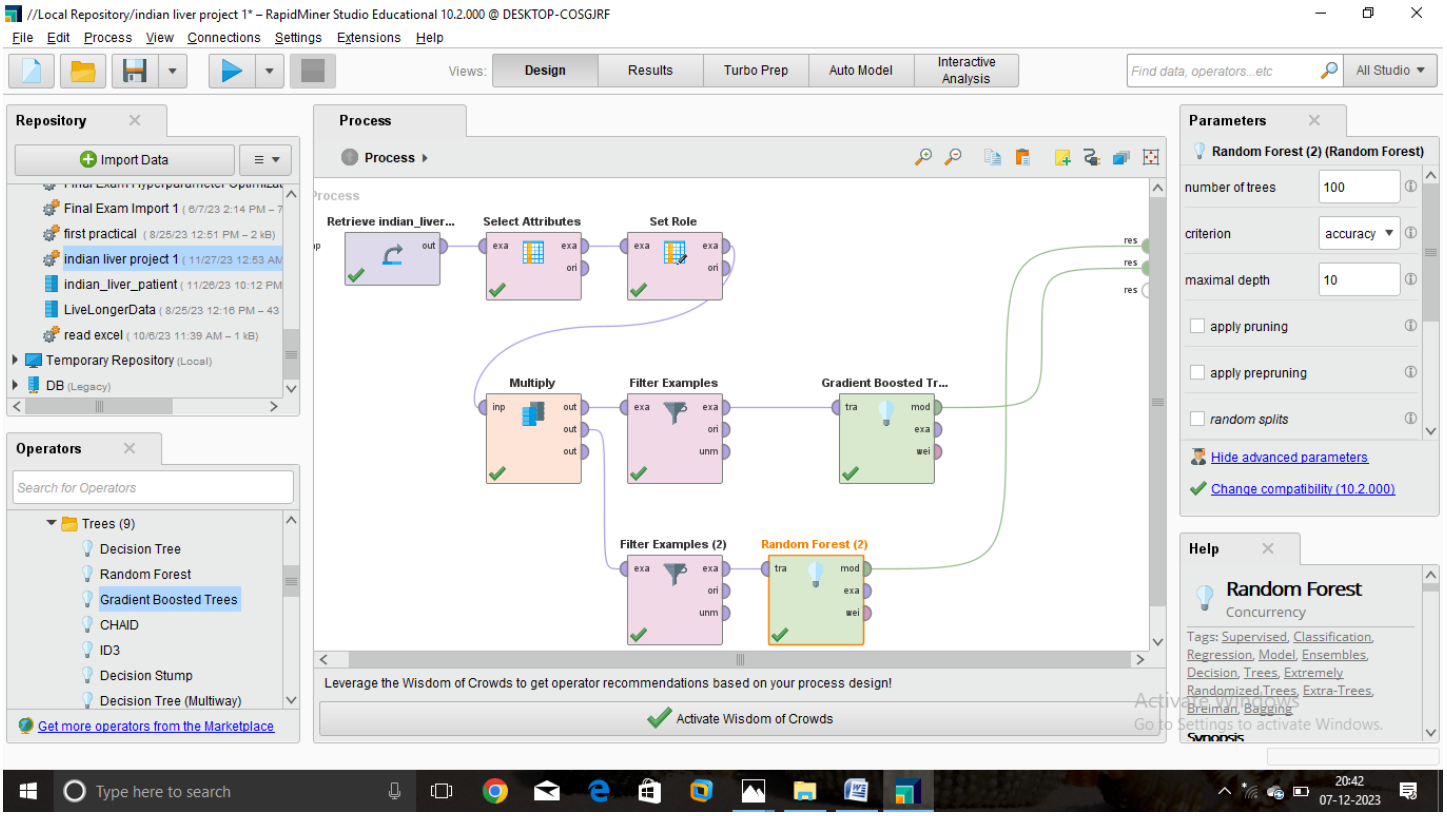


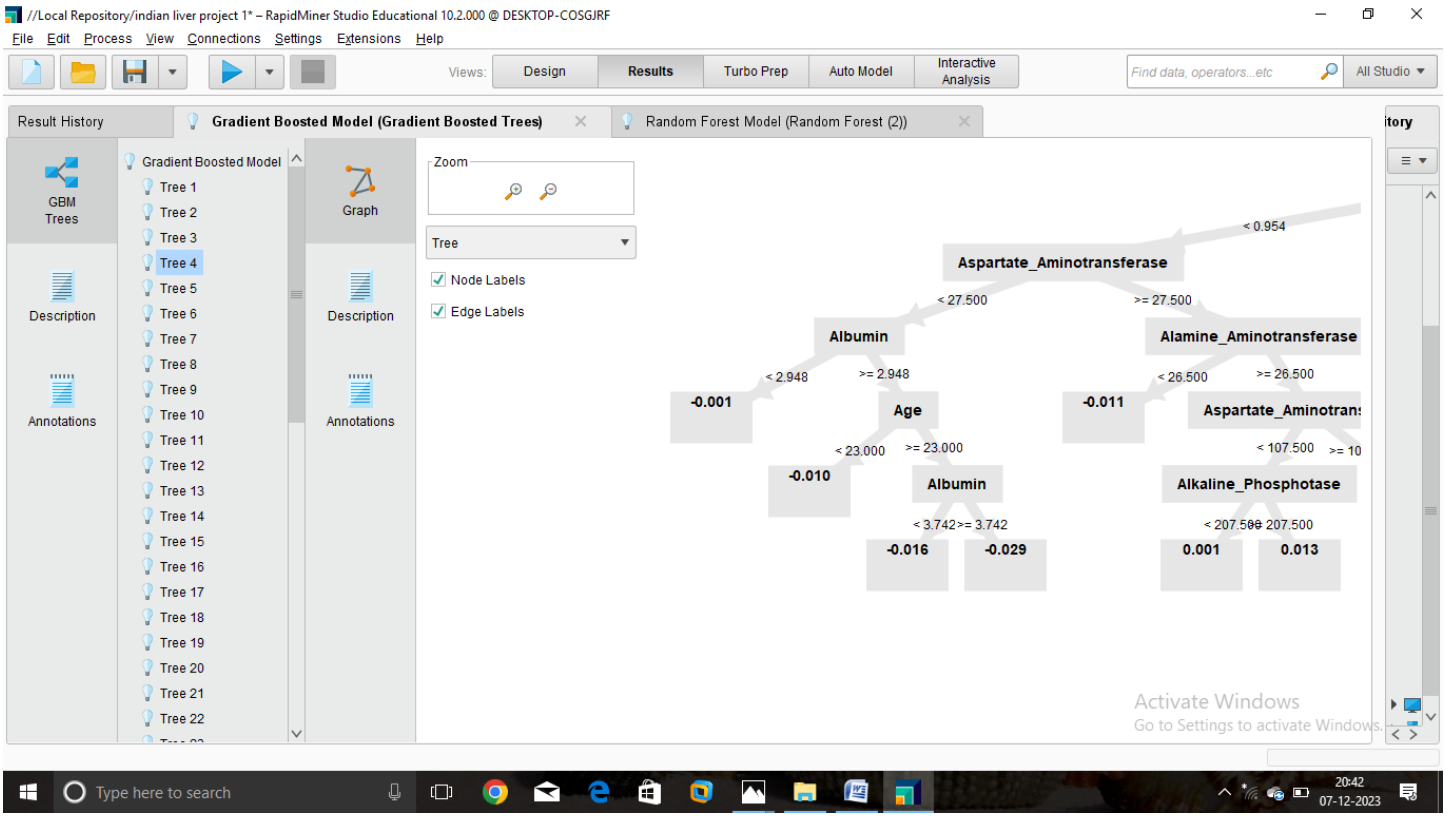
LIVER DISEASE PRIDICTION IN DIFFERENT AGE GROUPS





LIVER DISEASE PRIDICTION IN DIFFERENT AGE GROUPS





Background

In this section, we first describe the dataset used for this study. Then, the basic concepts of data mining and its methods are discussed. Thereafter, we illustrate the classification models used in this study. Lastly, we briefly explain the feature selection methods in data mining.

Liver disease dataset description

In this study, the liver disease data were collected from the UCI Machine learning repository . The dataset contains 584 records with 11 features including Age, Gender, TB, DB, Alkphos, Sgpt, Sgot, TP, ALB, A/G ratio, and target label. provides a detailed description and the type of features. The dataset was split into two sets that included 416 records for group 1(liver patients) and 167 records for group 2(non-liver patients).

Data Mining

Applying data mining models in medical research such as for Liver Disease is a significant undertaking because of the large volumes of available Liver-related datasets for extracting knowledge. Hence, it is necessary to provide a general schema of data mining by presenting a tree structure. Data mining methods are mainly divided into two categories: descriptive and predictive methods.

The definitions of basic models

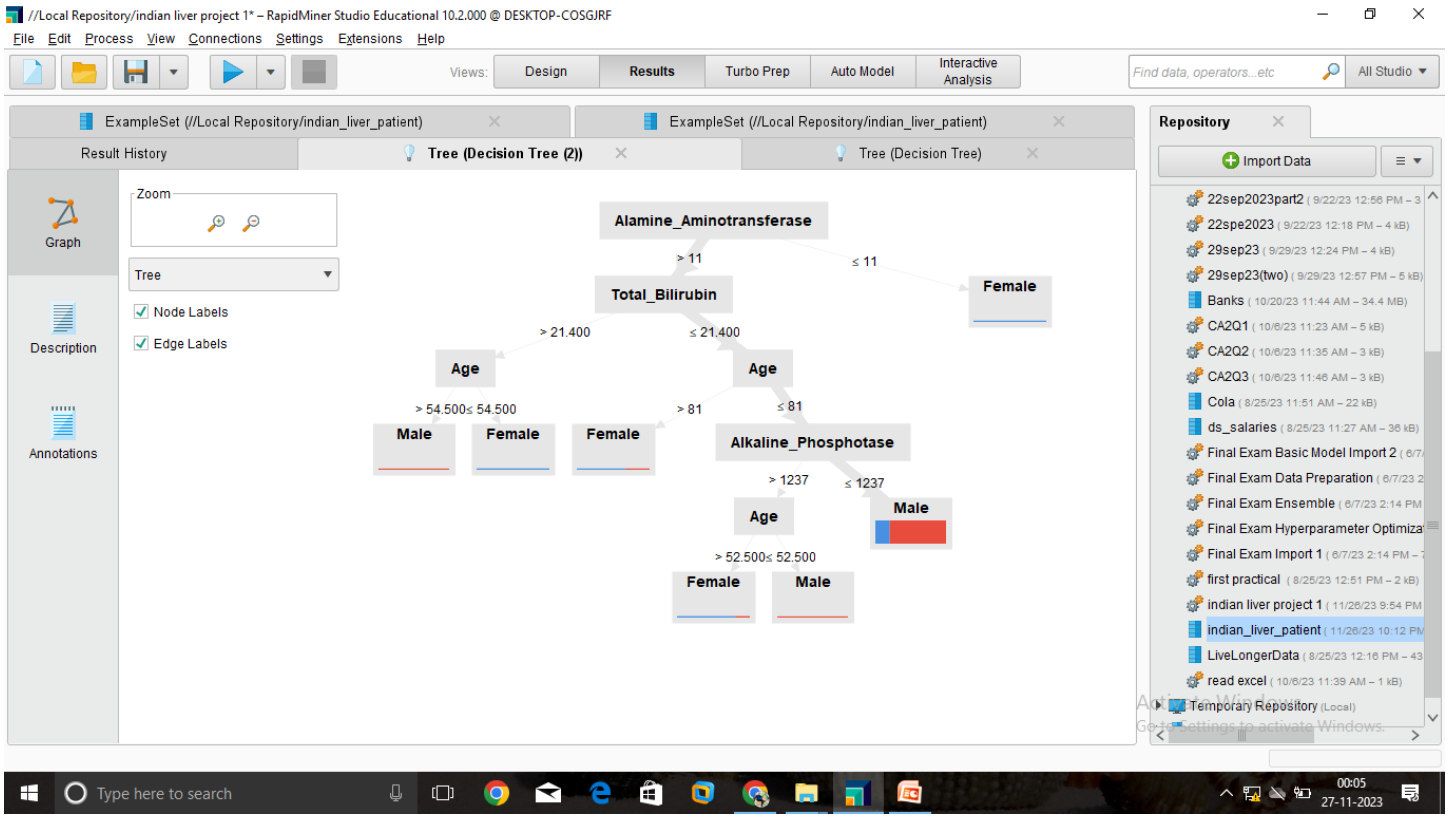
2.3.1. Random forest

The random forest model is one of the descriptive models of machine learning, and is a competitive predictive application in various fields such as medicine, finance, chemical engineering, and so on.

To improve the efficiency of a random forest model one can increase accuracy and speed . The model generates several trees, and chooses the most significant votes. In the field of improving accuracy, it uses the evaluation of various features and combines

functions. To classify, it puts each input vector to each of the trees in the forest. The random forest model on UCI liver disease dataset.

We have used decision tree model for our prediction model
Where the model will predict the chances of liver disease in 2 different age groups i.e
i)Above 45 years
ii)Below 45 years



Experimental results

In this section, the results of the classification models obtained by using the confusion matrix on Seven criteria.

By comparing the performance of data mining models, the accuracy of the Random forest, MLP Neural network, Bayesian network, SVM models was 86.26%, 78.11%, 66.09%, and 75.10%, respectively, while the accuracy of the PSO-SVM model is 94.42% based on the ELTA approach.

According to the other criteria, the method of PSO-SVM had the highest sensitivity, specificity, precision, and F-measure, and is the best predictive model. The obtained results of evaluation of criterion for models based on the ELTA approach are demonstrated.

Also, in this study, using a 10-fold cross-validation method, our models were evaluated on the liver disease dataset, with an average estimated accuracy of classification models calculated as 87.35%, 78.91%, 66.78%, 76.51%, and 95.17% for Random forest, MLP Neural network, Bayesian network, SVM and PSO-SVM models, respectively.

Table 3. The obtained results of evaluation of criterion for models based on the ELTA approach.

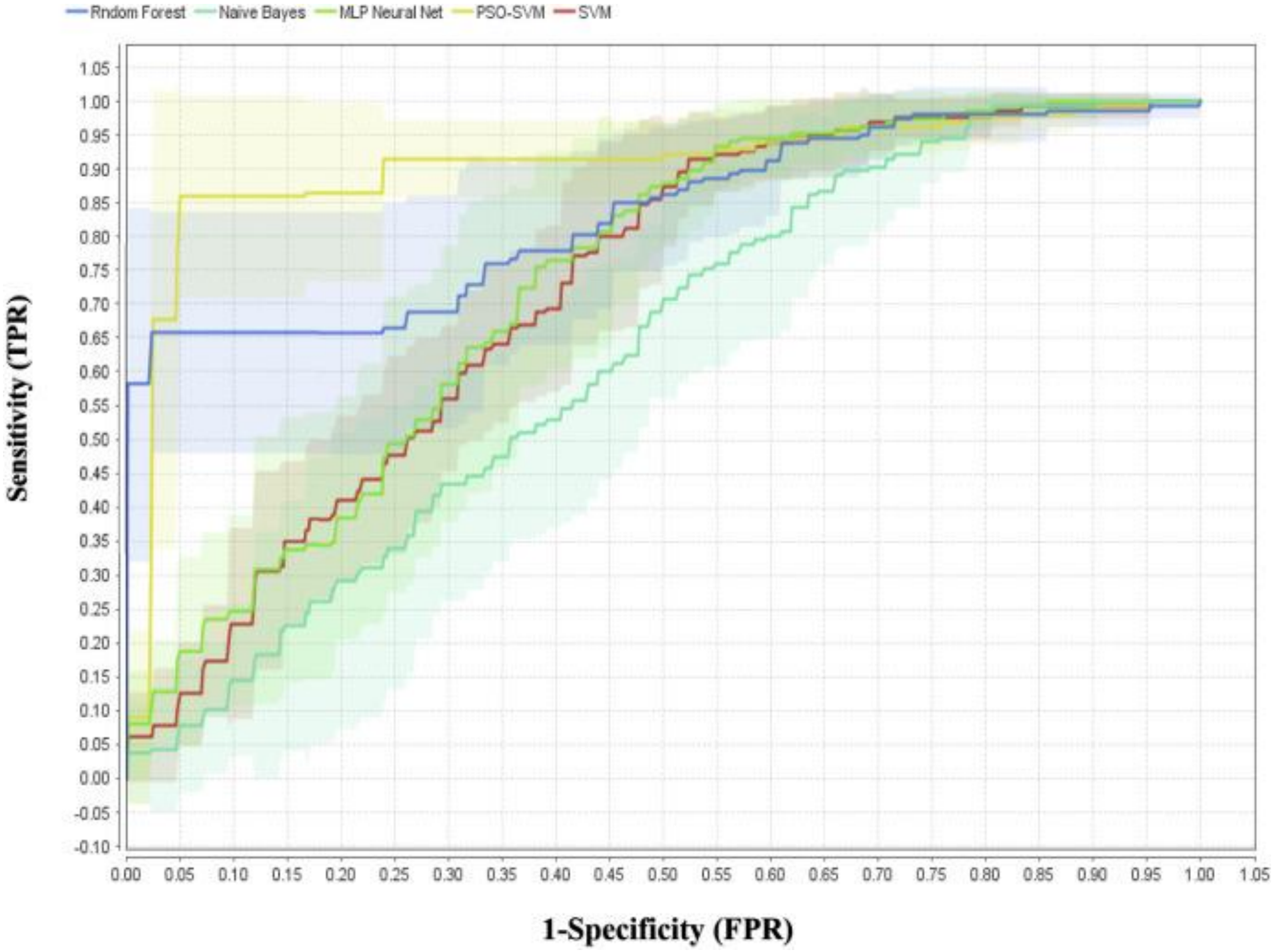
(Reference from https://www.sciencedirect.com/science/article)

Classification models	Accuracy	Sensitivity	Specificity	Precision	F-measure	FPR
Random Forest	86.26	87.65	83.09	92.20	89.86	16.91
SVM	75.10	81.59	60.00	82.60	82.09	40.00
Naïve Bayes	66.09	75.46	44.28	75.92	75.68	55.72
MLP Neural Network	78.11	82.53	67.16	86.16	84.30	32.84
PSO-SVM	94.42	94.93	93.33	96.77	95.84	6.67

Furthermore, another significant criterion utilized to determine the efficiency of a classification model is the AUC criterion. The AUC represents the surface area below the graph ROC.

The ROC is a two-dimensional graphical diagram illustration of the trade-off between the true positive rate (sensitivity) and false positive rate (1-specificity). The larger AUC value represents the higher performance of the model. The ROC curve was demonstrated for the models in

LIVER DISEASE PRIDICTION IN DIFFERENT AGE GROUPS



Hence, confidence for group 1, for example in scope 0.31 related to the seventh record includes 142 records demonstrating that 53 persons suffered from the disease. Therefore, confidence for group 1 on datasets at the scope 0.31 is very significant for 142 records, showing that over 95% of patients suffered from the disease.

Results and discussion

We have exploited process data mining from the ELTA approach. This approach includes two sections of preprocessing (ELT, (Extraction, Loading, Transformation)) and analysis (A).

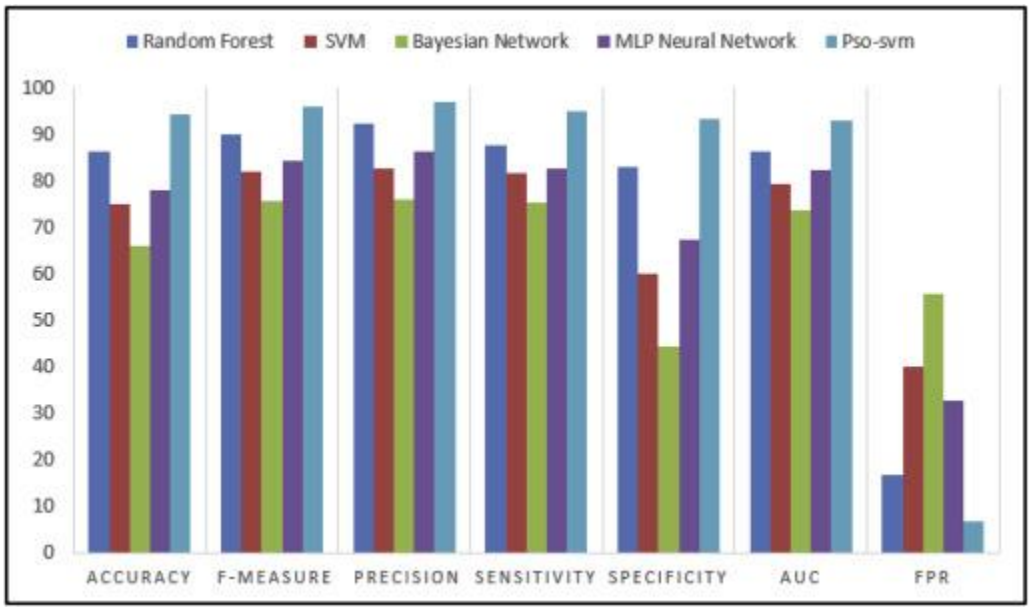
The best result was obtained by combining PSO-SVM for improving performance. According to ELTA, classification models were compared in terms of seven criteria - accuracy, Sensitivity or Recall, Specificity, precision, F-measure, FPR, and AUC. The main purpose of this study was to compare multiple prediction models for liver disease by selecting the significant features based on the ELTA approach.

This study developed five widely-used data mining classification models, Random Forest, MLP-Neural Network, Bayesian network, SVM, and PSO-SVM, along with a 10-fold cross-validation method. Using this method, our models were evaluated so that the average of estimated accuracy was calculated, found to be 87.35%, 78.91%, 66.78%, 76.51% and 95.17% for the Random forest, MLP Neural network, Bayesian network, SVM, and PSO-SVM models, respectively.

Regarding , the PSO-SVM model had the best performance in terms of the aforementioned criteria. Furthermore, one of the significant sections of this study was the selection of features for predicting liver disease comparing different models regarding the selection of these features based on .

Finally, the PSO-SVM model was best with selection of seven features and voting operation, in which Alkphos with the rate of 0.75 was the most significant feature, and age with the rate of 0.34 was the least significant feature.

In addition, the Lift Chart diagram for the model of PSO-SVM was created on the test dataset with confidence index, based on the liver disease, for the target group in the records being very significant for the diagnosis of healthy people versus patients. The proposed model demonstrated better performance in terms of accuracy, f-measure, precision, sensitivity, specificity, AUC, and FPR criteria.



Conclusion

In the present study, a systematic effort was done for Medical Data Mining of Liver disease on the UCI dataset.

Timely prediction of Liver disease is significant especially regarding its accuracy. Since, the main purpose of this paper was to select the most significant feature in achieving the highest accuracy in predicting liver disease according to ELTA approach, the authors examined five classification models, including Random Forest, MLP-Neural network, Bayesian network, SVM, and PSO-SVM. Ultimately, using a PSO-SVM model, seven features were extracted, with the highest accuracy.

Furthermore, another purpose of this paper was to compare the performance of the models mentioned in terms of accuracy, sensitivity, specificity, AUC, F-measure, precision, and FPR criteria. After comparing them, the results demonstrated that the PSO-SVM model had the best performance compared to other models Moreover, a 10-fold cross validation method for evaluation of models was used. Using this method, the PSO-SVM model had the highest average of estimated accuracy as compared to other classification models.

As future work, the PSO-SVM model can be used in a real laboratory environment. Consequently, these findings can be useful as a suitable and conducive way to identify people with liver disease or without liver disease in the real environment. Meta-Heuristic models could be used for optimizing the classification models, such as the Ant Colony System (ACS) and Genetic Model (GM) in the field of Liver disease.

In addition, in the application of smart algorithms in the diagnosis and prediction of diseases, in particular liver disease, deep neural networks implemented on the liver disease dataset and other datasets can be used to improve accuracy of detection, by selecting more effective and accurate features.

Refrences:

REFERENCES [1] M. Sameer and B. Gupta, “Beta Band as a Biomarker for Classification between Interictal and Ictal States of Epileptical Patients,” in 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN), 2020, pp. 567–570, doi: 10.1109/SPIN48934.2020.9071343. [2] S. K. B. Sangeetha, N.

Afreen, and G. Ahmad, “A Combined Image Segmentation and Classification Approach for COVID-19 Infected Lungs,” J. homepage http//iieta. org/journals/rces, vol. 8, no. 3, pp. 71–76, 2021. [3] M. Sameer, A. K. Gupta, C. Chakraborty, and B. Gupta, “Epileptical Seizure Detection: Performance analysis of gamma band in EEG signal Using Short-Time Fourier Transform,” in 2019 22nd International Symposium on Wireless Personal Multimedia Communications (WPMC), 2019, pp. 1–6, doi: 10.1109/WPMC48795.2019.9096119. [4] A. Mahajan, K. Somaraj, and M. Sameer, “Adopting Artificial Intelligence Powered ConvNet To Detect Epileptic Seizures,” in 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2021, pp. 427–432, doi: 10.1109/IECBES48179.2021.9398832. [5] N. Nasir, N. Afreen, R. Patel, S. Kaur, and M. Sameer, “A Transfer Learning Approach for Diabetic Retinopathy and Diabetic Macular Edema Severity Grading,” Rev. d’Intelligence Artif., vol. 35, pp. 497–502, Dec. 2021, doi: 10.18280/ria.350608. [6] M. Sameer and B. Gupta, “ROC Analysis of EEG Subbands for Epileptic Seizure Detection using Naive Bayes Classifier,” J. Mob. Multimed., pp. 299–310, 2021. [7] M. Sameer and B. Gupta, “Time–Frequency Statistical Features of Delta Band for Detection of Epileptic Seizures,” Wirel. Pers. Commun., 2021, doi: 10.1007/s11277-021-08909-y. [8] S. M. Beeraka, A. Kumar, M. Sameer, S. Ghosh, and B. Gupta, “Accuracy Enhancement of Epileptic Seizure Detection: A Deep Learning Approach with Hardware Realization of STFT,” Circuits, Syst. Signal Process., 2021, doi: 10.1007/s00034-021- 01789-4. [9] S. Gupta, M. Sameer, and N. Mohan, “Detection of Epileptic Seizures using Convolutional Neural Network,” in 2021 International Conference on Emerging Smart Computing and

Informatics (ESCI), 2021, pp. 786–790, doi: 10.1109/ESCI50559.2021.9396983. [10] P. Porwal et al., “Indian Diabetic Retinopathy Image Dataset (IDRiD): A Database for Diabetic Retinopathy Screening Research,” Data , vol. 3, no. 3. 2018, doi: 10.3390/data3030025. [11] M. Sameer and P. Agarwal, “Coplanar waveguide microwave sensor for label-free real-time glucose detection,” Radioengineering, vol. 28, no. 2, p. 491, 2019. [12] M. Sameer and B. Gupta, “Detection of epileptical seizures based on alpha band statistical features,” Wirel. Pers. Commun., vol. 115, no. 2, pp. 909–925, 2020, doi: 10.1007/s11277-020-07542-5. [13] M. Sameer, A. K. Gupta, C. Chakraborty, and B. Gupta, “ROC Analysis for detection of Epileptical Seizures using Haralick features of Gamma band,” in 2020 National Conference on Authorized licensed use limited to: University of the Cumberlands. Downloaded on June 08,2022 at 23:48:44 UTC from IEEE Xplore. Restrictions apply. 226 Communications (NCC), 2020, pp. 1–5, doi: 10.1109/NCC48643.2020.9056027. [14] N. Afreen, R. Patel, M. Ahmed, and M. Sameer, “A Novel Machine Learning Approach Using Boosting Algorithm for Liver Disease Classification,” in 2021 5th International Conference on Information Systems and Computer Networks (ISCON), 2021, pp. 1–5. [15] N. Jiwani, K. Gupta, and P. Whig, “Novel HealthCare Framework for Cardiac Arrest With the Application of AI Using ANN,” in 2021 5th International Conference on Information Systems and Computer Networks (ISCON), 2021, pp. 1–5, doi: 10.1109/ISCON52037.2021.9702493. [16] M. Sameer and B. Gupta, “CNN based framework for detection of epileptic seizures,” Multimed. Tools Appl., 2022, doi: 10.1007/s11042-022-12702-9. [17] G.S. Tomar, S. Verma & Ashish Jha; “Web Page

Classification using Modified naïve Baysian Approach", IEEE TENCON-2006, pp 1-4,

14-17 Nov 2006