Chettinad Health City Medical Journal
Volume 13, Issue 4 - 2024, Pg. No. 21-33
Peer Reviewed & Open Access Journal

Research Article

An Efficient Hybrid Data Mining Model for Prognostication of an Imbalanced Data Set of Liver Disorder: A K-Prototype Naïve Bayes Approach

Divya[1], Vineeta Singh[2], Ravins Dohare[3], Manoj Kumar[4,5]

[1]Research Scholar, [2]Professor, Department of Statistics, Institute of Social Sciences, Dr Bhimrao Ambedkar University, Agra, India.
[3]Professor, Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, Jamia Nagar, New Delhi, India.
[4]Assistant Professor, Centre for Economic Studies and Planning, Jawaharlal Nehru University, India.
[5]Postdoctoral Associate, McGowan Institute for Regenerative Medicine (MIRM), Department of Surgery, University of Pittsburgh, Pittsburgh, PA, USA.
DOI: https://doi.org/10.24321/2278.2044.202456

I N F O

Corresponding Author:
Divya, Department of Statistics, Institute of Social Sciences, Dr Bhimrao Ambedkar University, Agra, India.
E-mail Id:
divyascholar7999@gmail.com
Orcid Id:
https://orcid.org/0000-0002-6381-6707
How to cite this article:
Divya, Singh V, Dohare R, Kumar M. An Efficient Hybrid Data Mining Model for Prognostication of an Imbalanced Data Set of Liver Disorder: A K-Prototype Naïve Bayes Approach. Chettinad Health City Med J. 2024;13(4):21-33.

Date of Submission: 2024-06-30
Date of Acceptance: 2024-09-09

A B S T R A C T

Background: Liver disorders have recently become the deadliest disorder in many countries, with the number of patients increasing as a result of alcohol consumption, exposure to toxic gases, and ingestion of tainted foods and drugs. Data mining is the most effective approach for detecting the disease early on.

Objective: This study aimed to predict and diagnose early-stage liver disorders.

Method: In this study, we used the Indian liver patient dataset from the UCI machine learning repository. This dataset contains the sex imbalance for which we applied both oversampling and undersampling strategies; we used principal component analysis (PCA) for feature selection. In this research, we built eight models from 4 experiments in RStudio with the required packages. These models are compared based on the performance factors, which include accuracy, sensitivity, specificity, and error rate. We constructed the Naïve Bayes model and a new innovative hybrid model combining k-prototype clustering and the Naïve Bayes classifier (K-PNB).

Results: The hybrid model gave a classification accuracy of 94%, a sensitivity of 99%, a specificity of 90% and a low error rate of 0.05%.

Conclusion: The findings showed that the proposed hybrid model (the K-PNB) outperformed the other models, which detect and diagnose liver disease in the early stages in very little time.

Keywords: Data Mining, PCA, K-Prototype Clustering, Naïve Bayes, Imbalanced Data, K-PNB

Introduction

The vast amount of data on liver patients is increasing daily, and the number of factors associated with liver disorders is continuously growing. The death rate from liver illness rose from 1.9% to 2.4% globally from 1990 to 2017. Around 400 million diabetics and 75 million people with alcohol use are at high risk of developing liver disease.[1] Alcohol consumption is a significant factor in causing liver cirrhosis. In India, liver disorders are a primary public health concern.[2] According to the most recent WHO report, liver disease mortality in India totalled 259,749 in 2017, accounting for 2.95% of fatalities and 18.3% of cirrhosis deaths globally.[3,4] This raises the burden of liver and other diseases and is associated with out-of-pocket expenditure to the population.[5] Therefore, it is difficult to predict and diagnose liver disorders in less time with an effective procedure. The present work is an attempt to achieve this goal. Data mining is a procedure for detecting and diagnosing a disease at an early stage. Many algorithms are very effective in predicting the risk of liver disorders. This work focused on building a hybrid model of unbalanced and balanced data by using data mining algorithms., i.e. principal component analysis (PCA) for feature selection, K-prototype clustering for clustering the mixed type of numeric and categorical data type, and a naïve Bayes model to classify whether the patients had the disease.

Ramana et al. compared liver patients from India and the USA.[6] In this study, they compared three common attributes with three experiments. Three experiments with seven combinations. In Experiment 1, they checked that there exists a more significant difference between all the possible attribute combinations, as in Experiment 2, Experiment 3 showed a more substantial difference between the attribute combinations except for SGPT between non-liver patients in the USA and India. Saxena, in 2013, used the K-medoids algorithm on primary and secondary datasets of liver patients.[7] The results show that this algorithm is more accurate, easily found, and has less competition time than K-means and the PAM algorithm, which also shows the best results. Mohan, 2015 compared the SVM and Naïve Bayes classifier based on performance factors such as classification accuracy and execution time.[8] The comparison showed that SVM is a better performer than the Naïve Bayes classifier. Roy et al. classified the liver disorders using SVM, LDA, diagonal linear discriminant analysis, quadratic discriminate analysis, diagonal QDA and Mahalanobis.[9] They compared their accuracy, sensitivity and specificity performance and found that SVM is the best performer, with the highest accuracy, 82% and the lowest error rate. 0.1701%. Baitharu & Pani developed the decision support system using machine learning algorithms such as Decision Tree, Naïve Bayes, ANN, ZeroR, 1BK (K- nearest neighbor) and VFI.[10] After comparing these algorithms, we found that ANN was the best classification algorithm, with an accuracy of 71% and an error rate of 0.3543%. Kuppan & Manoharan compared the machine learning algorithms such as Naïve Bayes, Decision Tree and J48.[11] The result showed that the decision tree is the best performer. Priya et al. detected liver disease in the early stage.[12] With the classification algorithm, random forest, SVM, J48, MLP and Bayesian Network and compared the accuracy of these algorithms. Hence, it found that SVM is the best classifier. Durai et al. compared the machine learning algorithm to predict liver disease.[13] They also found that the J48 algorithm achieved the highest prediction accuracy of 95.04%. Razali et al. predicted liver disease using classification algorithms such as Neural Networks and Naïve Bayes algorithm and compared results by classification, accuracy, preseason, recall and F-score. They found that Naïve Bayes is the best performer.[14] Yajurved et al. predicted liver disease using data mining models such as Random Forest, Logistic regression and SVM; they compared the model accuracy and found that Random Forest outperformed them.[15] Baiju et al. formed the hybrid model using a decision tree and Naïve Bayes algorithms. They evaluated their results from accuracy, precision, recall, F-measures, and ROC and achieved a classification accuracy of 98% satisfactory results.[16]

Method

Dataset Description

The dataset from the UCI machine learning repository is used for liver disorders. The specific Indian liver patient data set (ILPD) contains 583 instances, with 11 attributes for 416 liver and 167 non-liver patients. The data includes 441 males and 142 females. The target variable is whether the individual has been diagnosed with liver disease or non-liver disease. Table 1 shows a brief description of the dataset. The analysis was performed using RStudio version 4.3.2. The architecture of the proposed model is shown in Figure 1.

Data Preprocessing

The variable A/G ratio has four missing values. Figure 2 shows the overall summary of missing values and missing value patterns. Median imputation was used to address these four values. This strategy retains the maximum instances by replacing the missing data with a value determined from the existing data. After handling missing values, we went through several phases of data analysis, including correlation analysis, data visualisation, and configuring the data types. We assigned sex to 1 and 2 numerical types, the minimum-maximum scaler function was used for the normalisation of data, and the target variable was assigned to a binary variable with levels 1 for non-liver disease and 2 for liver disease.

ISSN: 2278-2044
DOI: https://doi.org/10.24321/2278.2044.202456

Correlation Analysis

Correlation analysis aims to determine the nature and degree of association between the predictors and response variables. The correlations between the variables and the response variable are shown in Figure 3. According to Table 2, the correlation between TB and DB is high at 0.875, and SGOT and SGPT are highly correlated at 0.792. ALB and TP were also highly correlated at 0.784, and the A/G ratio and ALB showed a moderate correlation of 0.686. Other variables also show a low positive and low negative correlation.

Addressing Class Imbalance

The original dataset revealed a significant target class imbalance of 167 healthy individuals and 416 disease patients and a sex imbalance of 142 (24%) females and 441 (76%) males. Similarly, to extend the models, we executed the ROSE package to mark these imbalances using the function ovun.sample. Both the oversampling and undersampling methods used minority class oversampling with replacement, and the majority class was undersampled without replacement. Table 2 shows the imbalance and balanced sex counts of the datasets.

Principal Component Analysis (PCA)

Principal component analysis (PCA) is a powerful method in several fields. This reduces the number of features while retaining most of the crucial information. It transforms new features called principal components, which account for the most significant proportion of variance in the original dataset. We used PCA for feature selection; regarding the effect of PCA on the dataset, we found that nine principal components and a total of six principal components explained almost 95.7% of the variance in the total variance in Experiment 3 and nearly 95.1% of the variance in the total variance in Experiment 4, as shown in Figure 4. So, we used the six principal components in Experiment 3 and Experiment 4. After PCA, the correlation chart shows the correlation between the transformed components and the response variable (TARGET) in Figure 5. According to it, the correlation between the components is zero, i.e., they are mutually independent.

Table 1.Description of the Attributes of Liver Disorder Dataset

S. No.	Attributes	Data Type	Description
1.	Sex	Binary	The gender of the patient (male and female)
2.	Age	Integer	Age of the patient. Any patient's age exceedings 89 is listed as "90".
3.	TB	Continuous	Total bilirubin
4.	DB	Continuous	Direct bilirubin
5.	Alkphos	Integer	Alkaline phosphatase
6.	SGPT	Integer	Alanine aminotransferase
7.	SGOT	Integer	Aspartate aminotransferase
8.	TP	Continuous	Total proteins
9.	ALB	Continuous	Albumin
10.	A/G ratio	Continuous	Albumin and globulin ratio
11.	TARGET	Binary	1 = Liver patient
			2 = Non-liver patient

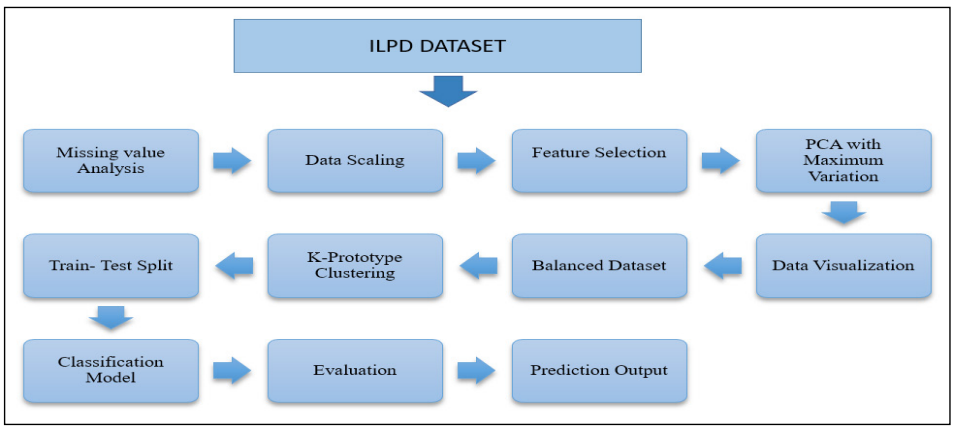

Figure 1.Architecture of the Proposed Model

ISSN: 2278-2044
DOI: https://doi.org/10.24321/2278.2044.202456

Table 2.Total Summary Counts of Classes in the Dataset, including Counts after the Dataset is Balanced

Gender	Target	Dataset.1 (Original Dataset) Sex		Dataset.2 (Oversampled Minority Class) Sex		Dataset.3 (Sex-Balanced,Oversampled Females) Sex	
		n (%)	N (%)	n (%)	N (%)	n (%)	N (%)
Female	1	92 (65)	142 (24)	377 (63)	601 (51)	261 (44)	595 (50)
	2	50 (35)		224 (37)		334 (56)	
Male	1	324 (73)	441 (76)	412 (70)	589 (49)	325 (55)	595 (50)
	2	117 (27)		177 (30)		270 (45)	
Total		583		1190		1190	

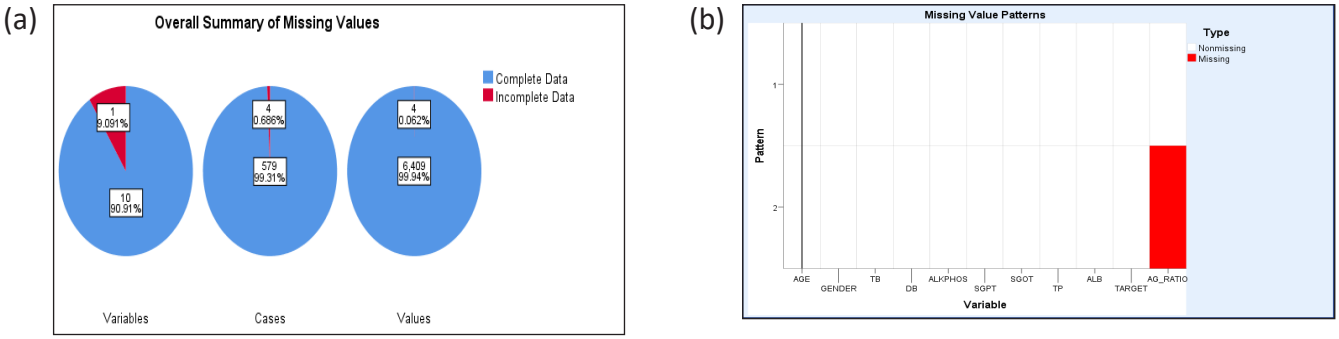

Figure 2(a).Overall Summary of Missing Value (b).Missing Value Pattern

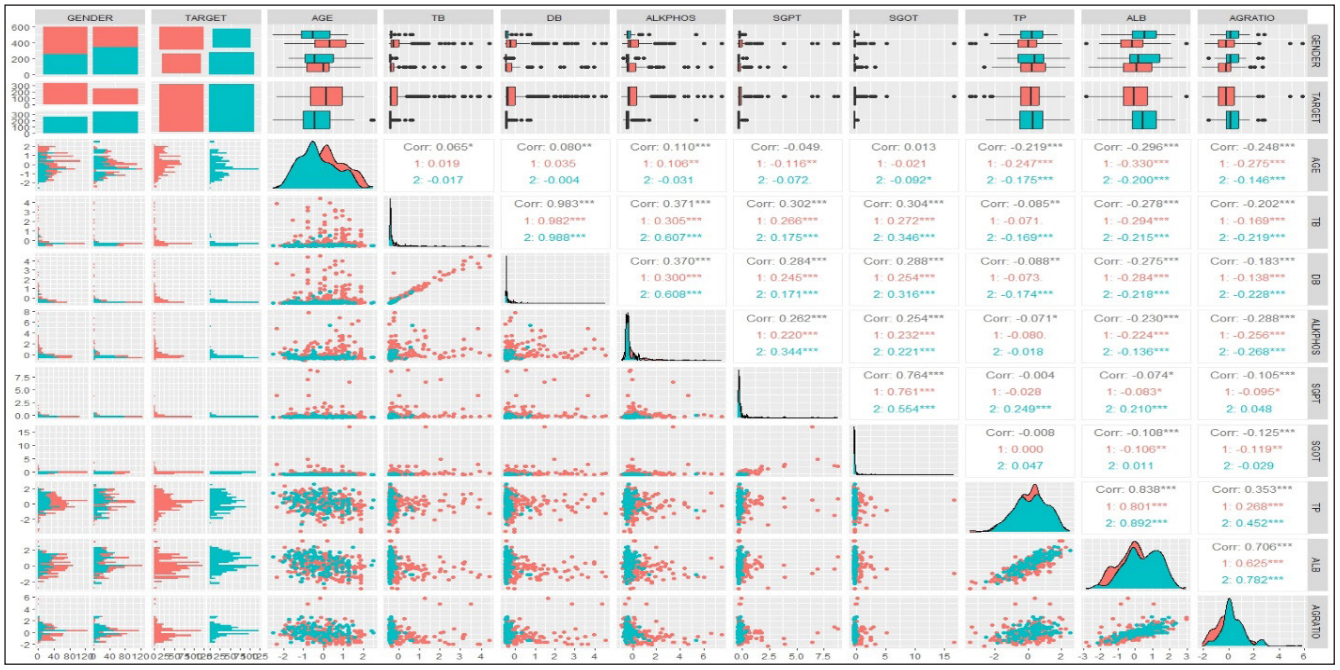

Figure 3.Correlation between the Predictors and Response Variable (TARGET)

ISSN: 2278-2044
DOI: https://doi.org/10.24321/2278.2044.202456

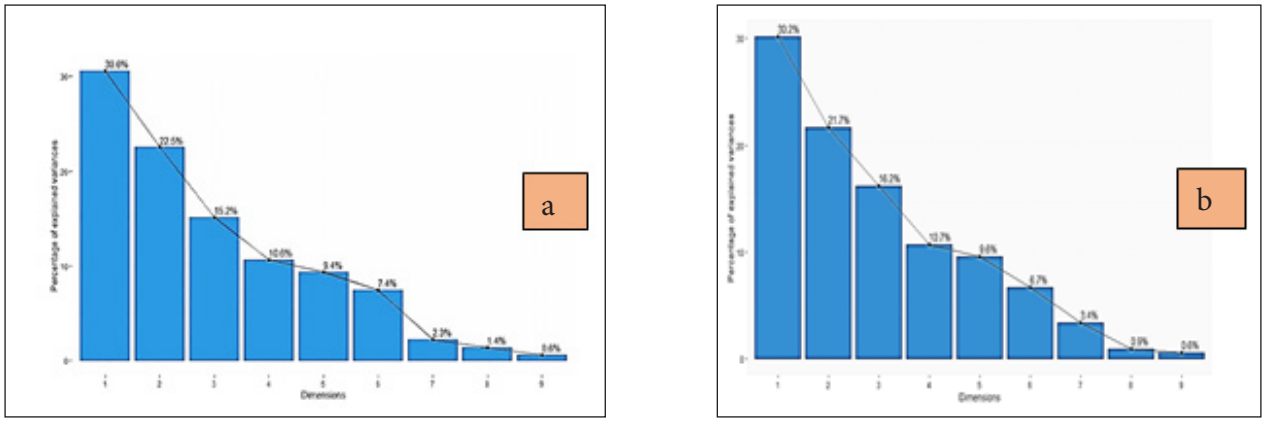

Figure 4(a). Scree Plot of Percentage of Explained Variation by Principal Components (b). Scree Plot of Percentage of Explained Variation by Principal Components

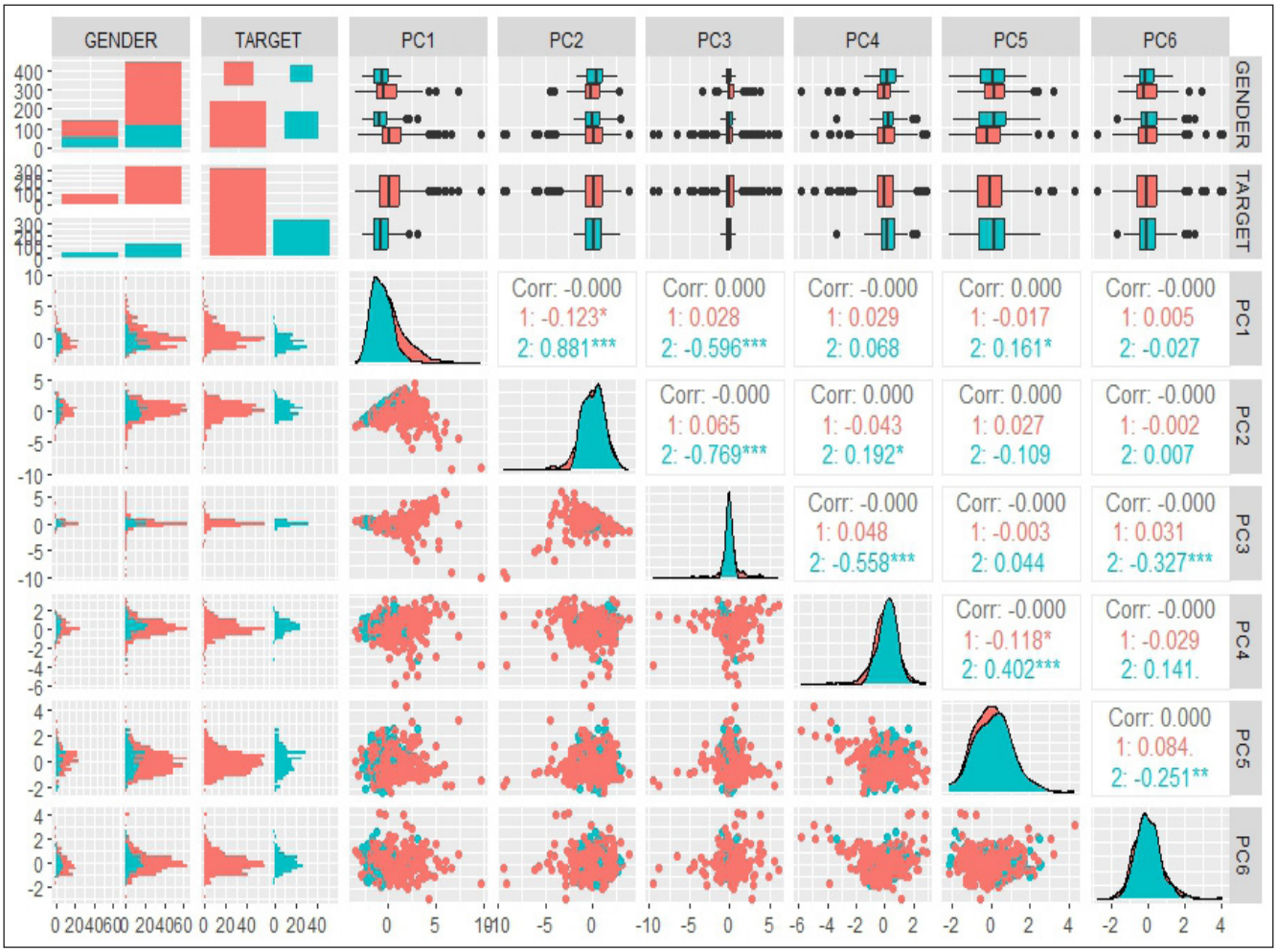

Figure 5.Correlation between the Transformed Variables and Response Variable (TARGET)

ISSN: 2278-2044
DOI: https://doi.org/10.24321/2278.2044.202456

K-Prototype Clustering

K-prototype clustering is an extension of the K-means and K-modes clustering algorithms.[17] It can handle both types of data, numerical and categorical. It can handle larger datasets to calculate the Euclidean distance for numerical data and measure the distance between categorical features using the number of matched categories. We used the clustMixType package for the k-prototype clustering algorithm.[18] Before implementing k-prototype clustering, we determined the optimal number of clusters using the function validation_kproto from the package clusMixType. In this function, we utilised the Silhouette approach, and the function kproto returned the results of k-prototype clustering, and function clprofiles shows the visualisation of k-prototype clustering result for cluster implementation, shown in Figures 6–9.

Naïve Bayes Classifier

The Naïve Bayes classifier is a supervised machine learning algorithm. It is based on the Bayes theorem, with the Naïve Bayes assumption of conditional independence between each pair of features (predictors) according to the value of the response variable. It is one of the simplest and most effective classifiers that assist in building a fast data mining classifier that can quickly make predictions. We employed the Naïve Bayes algorithm via the e1071 package with the naïveBayes function and prediction by the predict function.

Model Development

Experiment 1: Model Trained on the Imbalanced Dataset without Feature Selection

The first model was trained using an unbalanced dataset. We used the split.sample function from the caTools package to split the dataset into training and testing sets 80% and 20%, respectively. We employed the Naïve Bayes algorithm by the e1071 package with the naïveBayes function and prediction by the predict function. The model accuracy was determined using the Confusion matrix created by the caret package with the confusion matrix function. We made a hybrid model with K-prototype clustering and the Naïve Bayes algorithm and evaluated its performance by accuracy, sensitivity, specificity, and error rate.

Experiment 2: Model Trained on a Sex Balance Dataset without Feature Selection

In this experiment, we balanced our sex class by oversampling and undersampling and then repeated Experiment 1. We divided the data into training and testing sets (80% and 20%, respectively) and evaluated the model using evaluation matrices. Next, we created a hybrid model with k-prototype clustering and the Naïve Bayes method and evaluated its performance using evaluation matrices.

Experiment 3: Model Trained on an Imbalanced Dataset with Feature Selection

In this model, we applied principal component analysis (PCA) to identify features from the gender imbalance dataset, trained the Naïve Bayes model on 80% of the dataset and tested it on 20%. Then, we used K-prototype clustering and the Naïve Bayes algorithm to create a hybrid model and compared it to the evaluation matrices.

Experiment 4: Model Trained on a Sex Balance Dataset with Feature Selection

In this model, we used both oversampling and undersampling approaches and principal component analysis for feature selection and repeated Experiment 3. We trained the Naïve Bayes model on 80% of the data and tested it on 20%. Then, we created a hybrid model using K-prototype clustering and the Naïve Bayes classifier. We assessed the accuracy and evaluation matrices.

Results & Discussion

These studies aimed to examine the performance of the Naïve Bayes model and the hybrid K-prototype clustering Naïve Bayes model for liver disease datasets.

Table 3 displays the confusion matrix of the test dataset of Experiment 1 of the Naïve Bayes classifier, and Figure 6 shows a visualisation of the K-prototype clustering findings. In contrast, Table 4 displays the confusion matrix of the test dataset of the hybrid model utilising the K-prototype clustering and Naïve Bayes classifier.

From the confusion matrix of Experiment 1, we evaluated the accuracy, sensitivity, specificity, and error rate of both the Naïve Bayes and hybrid k-prototype clustering Naïve Bayes model. The optimal number of clusters is 2, the distance parameter lambda is 2.713815, the cluster sizes are 266 and 317, the within-cluster sums of the squares are 3161.583 and 1466.654, and the total within-cluster sum of a square is 4628.237.

Table 3.Confusion Matrix of Test Dataset of Experiment 1 of the Naïve Bayes Classifier

Prediction Class	Actual Class	
	1	2
1	40	0
2	43	33

ISSN: 2278-2044
DOI: https://doi.org/10.24321/2278.2044.202456

Table 4.Confusion Matrix of the Test Dataset of Experiment 1 of The Hybrid Model Utilising the K-Prototype Clustering and Naïve Bayes Classifier

Prediction Class	Actual Class	
	1	2
1	42	5
2	72	114

Table 5.Confusion Matrix of Test Dataset of Experiment 2 of the Naïve Bayes Classifier

Prediction Class	Actual Class	
	1	2
1	46	4
2	37	29

Table 6.Confusion Matrix of the Test Dataset of Experiment 2 of the Hybrid Model Utilising the K-Prototype Clustering and Naïve Bayes Classifier

Prediction Class	Actual Class	
	1	2
1	58	10
2	61	109

Table 7.Confusion Matrix of Test Dataset of Experiment 3 of the Naïve Bayes Classifier

Prediction Class	Actual Class	
	1	2
1	42	2
2	11	61

Table 8.Confusion Matrix of the Test Dataset of Experiment 3 of the Hybrid Model Utilising the K-Prototype Clustering and Naïve Bayes Classifier

Prediction Class	Actual Class	
	1	2
1	82	1
2	18	137

Table 9.Confusion Matrix of Test Dataset of Experiment 4 of the Naïve Bayes Classifier

Prediction Class	Actual Class	
	1	2
1	10	7
2	1	98

Table 10.Confusion Matrix of the Test Dataset of Experiment 4 of the Hybrid Model Utilising the K-Prototype Clustering and Naïve Bayes Classifier

Prediction Class	Actual Class	
	1	2
1	126	11
2	1	100

Table 5 displays the confusion matrix of the test dataset of Experiment 2 of the Naïve Bayes classifier, while Figure 7 visualises the K-prototype clustering findings. Table 6 displays the confusion matrix of the test dataset for the hybrid model that combines K-prototype clustering and Naïve Bayes. Classifiers. Further analysis of the results was carried out using the same packages and functions.

The confusion matrix from Experiment 2 assessed the accuracy, sensitivity, specificity, and error rate of both Naïve. Bayes and the hybrid model k-prototype clustering Naïve Bayes. The best number of clusters is 2, the distance parameter lambda is 1.466958, the cluster sizes are 501 and 689, the within-cluster sums of the squares are 4065.886 and 2612.690, and the total within-cluster sum of the squares is 6678.577.

Table 7 shows the confusion matrix of the test dataset of Experiment 3 of the Naïve Bayes classifier, Figure 8 shows the visualisation of the K-prototype clustering results, and Table 8 shows the confusion matrix of the hybrid model using K-prototype clustering and Naïve Bayes classifiers.

The confusion matrix from Experiment 3 was used to assess the accuracy, sensitivity, specificity, and error rate of both Naïve Bayes and the hybrid model k-prototype clustering of Naïve Bayes. The best number of clusters is 2, the distance parameter lambda is 3.896415, the cluster sizes are 57 and 526, the within-cluster sum of squares is 1306.245 and 3222.033, and the within-cluster sum is 4528.278.

Table 9 shows the confusion matrix of Experiment 4 of the Naïve Bayes classifier, Figure 9 shows the visualisation of the K-prototype clustering results, and Table 10 shows the confusion matrix of the hybrid model using K-prototype clustering and Naïve Bayes classifiers. The confusion matrix from Experiment 4 was used to assess the accuracy, sensitivity, specificity, and error rate of Naïve Bayes and the hybrid model k-prototype clustering of Naïve Bayes. The best number of clusters is 2, the distance parameter lambda is 2.361178, the cluster sizes are 633 and 557, the within-cluster sums of the squares are 2585.519 and 4873.853, and the total within-cluster sum of the squares is 7459.372. A comparison of the proposed models is shown in Table 11.

ISSN: 2278-2044
DOI: https://doi.org/10.24321/2278.2044.202456

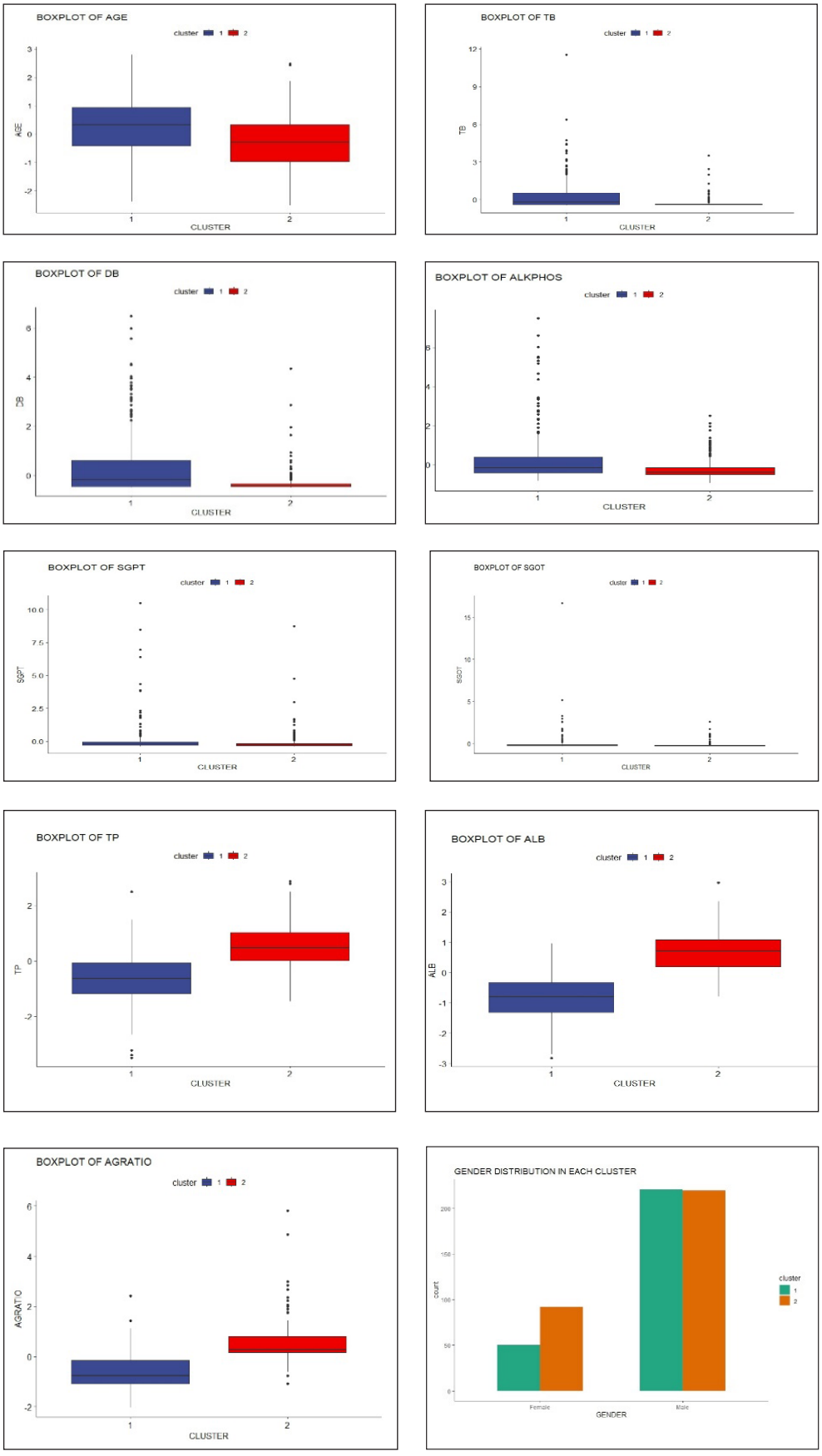

Figure 6.Visualisation of K-Prototype Clustering Results of Experiment 1

ISSN: 2278-2044
DOI: https://doi.org/10.24321/2278.2044.202456

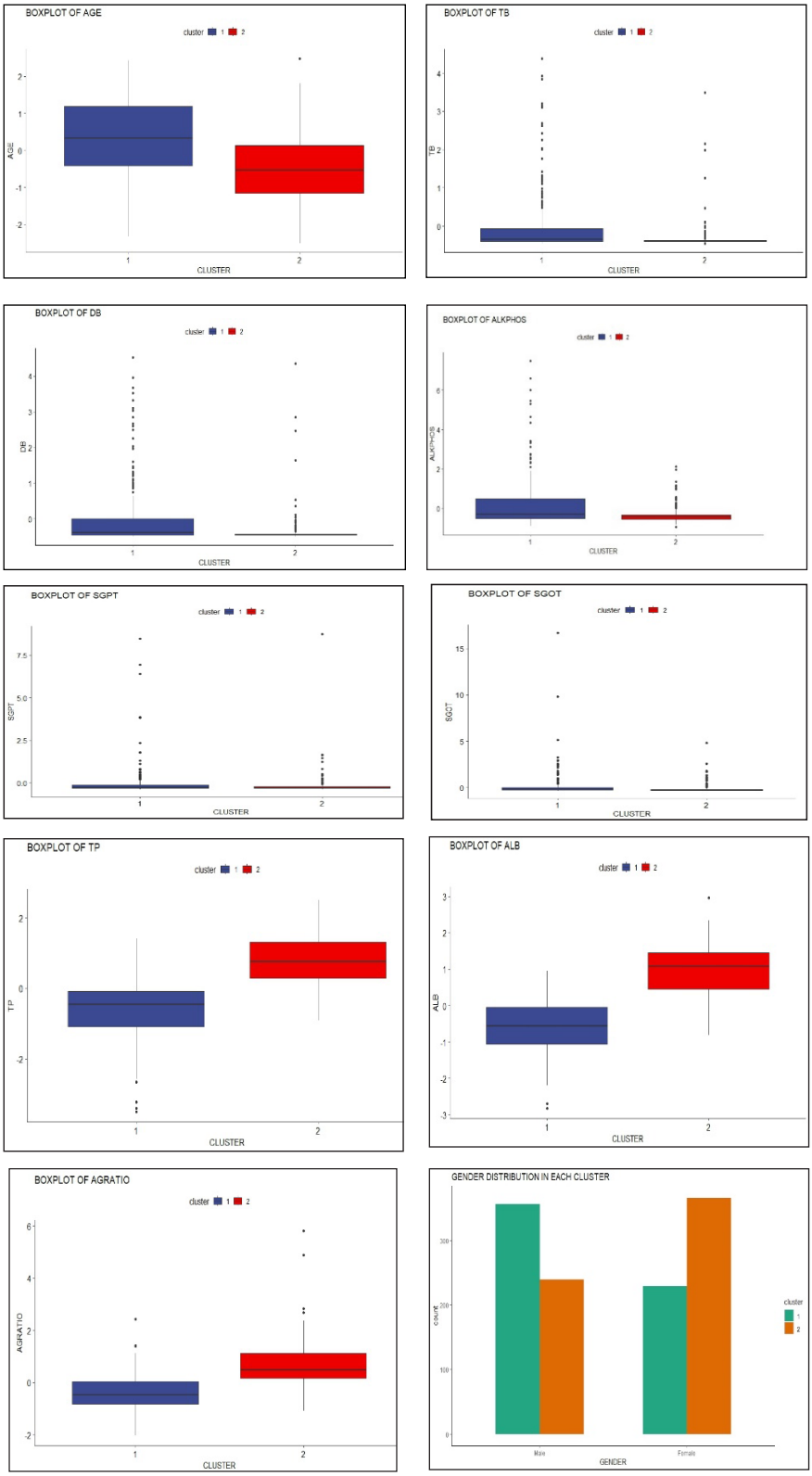

Figure 7.Visualisation of K-Prototype Clustering Results of Experiment 2

ISSN: 2278-2044
DOI: https://doi.org/10.24321/2278.2044.202456

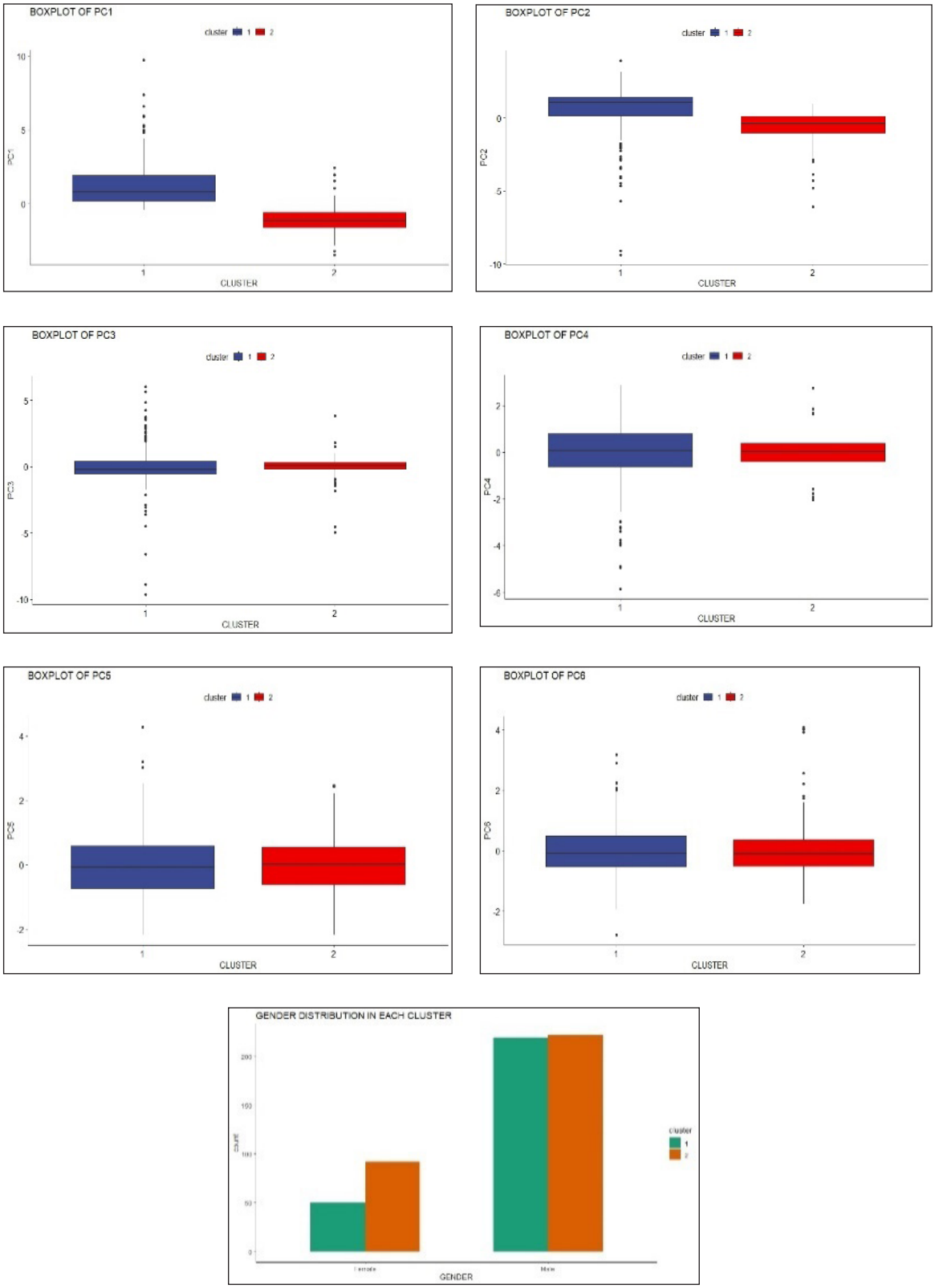

Figure 8.Visualisation of K-Prototype Clustering Results of Experiment 3

ISSN: 2278-2044
DOI: https://doi.org/10.24321/2278.2044.202456

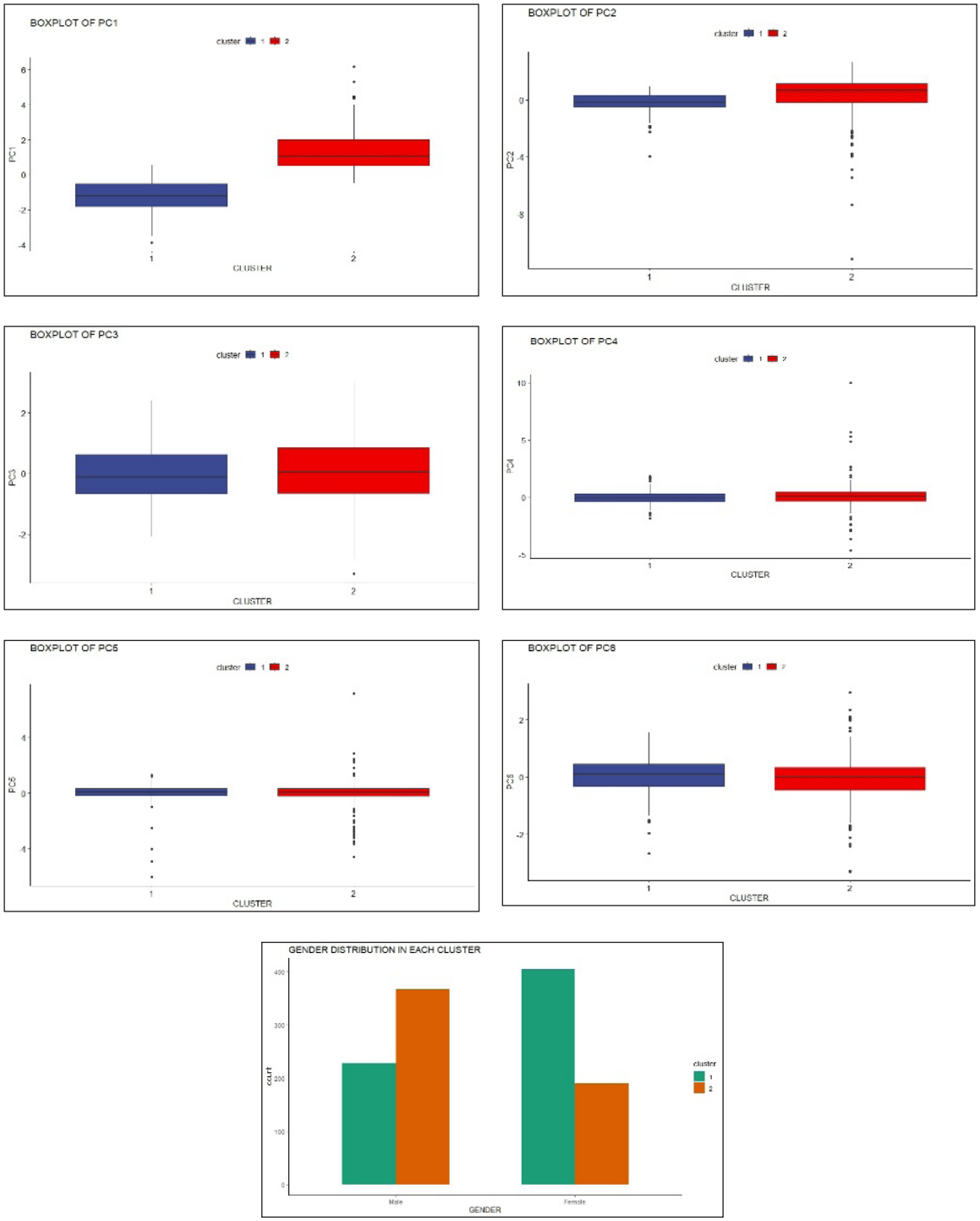

Figure 9.Visualisation of the K-Prototype Clustering Results of Experiment 4

ISSN: 2278-2044
DOI: https://doi.org/10.24321/2278.2044.202456

Table 11.Overall Comparison of the Proposed Models

Models	Accuracy (%)	Sensitivity (%)	Specificity (%)	Error Rate (%)
Experiment 1				
Naïve Bayes	62	48	100	0.370
K-prototype + Naïve Bayes	88	79	96	0.112
Experiment 2				
Naïve Bayes	67	39	95	0.330
K-prototype + Naïve Bayes	92	82	99	0.079
Experiment 3				
Naïve Bayes	64	55	87	0.080
K-prototype + Naïve Bayes	93	90	93	0.068
Experiment 4				
Naïve Bayes	70	48	91	0.250
K-prototype + Naïve Bayes	94	99	90	0.050

Conclusion

Data mining has significant importance in the healthcare industry. Clustering and classification are the most important data mining techniques for predicting and diagnosing diseases in the healthcare industry. This research introduces an innovative hybrid data mining model for predicting liver disease using the K-prototype and Naïve Bayes algorithm of clustering and classification techniques. From the experimental results, this work concludes that K-PNB is the best algorithm because it achieves the highest accuracy of 94%, a sensitivity of 99%, a specificity of 90% and a low error rate of 0.05%.

Future directions for improving liver disease prediction and classification include incorporating more diverse data sources and combining data mining techniques and training models to predict individual risk based on unique attributes. Developing explainable models is crucial for utilising machine learning to predict and classify liver diseases. Models should provide transparent and interpretable insights into the causes of liver disease. Explainable models improve healthcare workers' decision-making and patient care.

Source of Funding: None

Conflict of Interest: None

References

1. Asrani SK, Devarbhavi H, Eaton J, Kamath PS. The burden of liver diseases in the world. J Hepatol. 2019;70(1):151-71. [PubMed] [Google Scholar]
2. Roerecke M, Vafaei A, Hasan OS, Chrystoja BR, Cruz M, Lee R, Neuman MG, Rehm J. Alcohol consumption and risk of liver cirrhosis: a systematic review and meta-analysis. Am J Gastroenterol. 2019;114(10):1574-86. [PubMed] [Google Scholar]
3. Mondal D, Das K, Chowdhury A. Epidemiology of liver diseases in India. Clin Liver Dis (Hoboken). 2022;19(3):114-7. [PubMed] [Google Scholar]
4. Wu XN, Xue F, Zhang N, Zhang W, Hou JJ, Lv Y, Xiang JX, Zhang XF. Global burden of liver cirrhosis and other chronic liver diseases caused by specific etiologies from 1990 to 2019. BMC Public Health. 2024;24(1):363. [PubMed] [Google Scholar]
5. Goyanka R, Yadav J, Kumar M, Sagar SK. Utilisation and out-of-pocket expenditure for AYUSH outpatient care among older adults in India. Chettinad Health City Med J. 2023;12(1):54-64. [Google Scholar]
6. Ramana BV, Babu MS, Venkateswarlu NB. A critical comparative study of liver patients from USA and India: an exploratory analysis. Int J Comput Sci Iss. 2012;9(3):506-16. [Google Scholar]
7. Saxena P. Evolving efficient clustering patterns in liver patient data through data mining techniques. Int J Comput Appl. 2013;66(16):23-8. [Google Scholar]
8. Vijayarani S, Dhayanand S. Liver disease prediction using SVM and Naïve Bayes Algorithms. Int J Sci Engi Techno Res. 2015;4(4):816-20. [Google Scholar]
9. Roy S, Singh A, Shadev SK. Machine learning method for classification of liver disorders. Far East J Electron Commun. 2016;16(4):789-800.
10. Baitharu TR, Pani SK. Analysis of data mining techniques for healthcare decision support system using liver disorder dataset. Procedia Comput Sci. 2016;85:862-70. [Google Scholar]
11. Kuppan P, Manoharan N. A tentative analysis of liver disorder using data mining algorithms J48, decision table and Naive Bayes. Int J Comput Algor. 2017;6(1):37-40. [Google Scholar]
12. Priya MB, Juliet PL, Tamilselvi PR. Performance analysis of liver disease prediction using machine learning

algorithms. Int Res J Eng Technol. 2018;5(1):206-11. [Google Scholar]
13. Durai V, Ramesh S, Kalthireddy D. Liver disease prediction using machine learning. Int J Adv Res Ideas Innov Technol. 2019;5(2):1584-8. [Google Scholar]
14. Razali N, Mustapha A, Wahab MH, Mostafa SA, Rostam SK. A data mining approach to prediction of liver diseases. J Phys Conf Ser. 2020;1529. [Google Scholar]
15. Yajurved J, Prasad PS, Umamaheswari KM. Analysis of chronic disease (liver) prediction using machine learning. J Posit School Psychol. 2022;6(4):5489-96. [Google Scholar]
16. Baiju BV, Kirubanantham P, Saranya S, Kumaresan A, Prakash G. Liver disease diagnosis and prediction by hybrid data mining approach. AIP Conf Proc. 2023;2523:020045. [Google Scholar]
17. Huang Z. Extensions to the k-means algorithm for clustering extensive data sets with categorical values. Data Min Knowl Discov. 1998;2:283-304. [Google Scholar]
18. Aschenbruck R, Szepannek G. Cluster validation for mixed-type data. Arch Data Sci. 2020;6(1):1-12. [Google Scholar]

ISSN: 2278-2044
DOI: https://doi.org/10.24321/2278.2044.202456