INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Comparative Analysis of ML Models with Selection Methods for Early Predictive Analytics of Sepsis in ICU

Vandana[1]*, **Rita Chhikara**[2]

1 Assistant Professor, The NorthCap University, India
2 HOD (CSE), The NorthCap University, India

OPEN ACCESS

Received: 10-06-2024
Accepted: 10-10-2024
Published: 18-11-2024

Citation: Vandana , Chhikara R (2024) Comparative Analysis of ML Models with Selection Methods for Early Predictive Analytics of Sepsis in ICU. Indian Journal of Science and Technology 17(41): 4338-4348. https://doi.org/10.17485/IJST/v17i41.1925

* **Corresponding author**.

tovandanamehta@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: This study aims to make early prediction of Sepsis using ML algorithms and provide comparative analysis of different feature selection techniques. **Methods:** In this study, the Physionet website dataset has been used. Sepsis categorization is done on the basis of Sepsis-3 criteria. The dataset used is highly imbalanced, with only 2% of the data belonging to Sepsis. The SmoteTomek technique is used to handle an imbalanced dataset. Various filter, embedded, and wrapper feature selection techniques, like tree-based feature selection technique, recursive feature elimination (RFE), information gain, Bhattacharya distance, lasso, etc., have been applied for the top-performing classification ML models. These selection techniques were applied to RandomForest, K Nearest Neighbors (KNN), and Decision Tree models. We compared the impact of these selection techniques on the aforesaid machine learning models. **Findings:** After applying the RFE technique, the Area Under the Receiver Operating Characteristic curve (AUROC) score of the RandomForest model has slightly increased from 0.996 to 0.9974. KNN model with a tree-based feature selection technique showed the highest sensitivity of 0.934. which is slightly higher than the sensitivity of 0.922, which was without applying any feature selection technique. Along with the AUROC score, the highest performance, in terms of specificity (0.9976), accuracy (0.9959), and f-measure score (0.9062), is achieved when RFE is applied to the RandomForest model. The best selection algorithm for decision trees and KNN is the tree-based selection technique. RFE is the best selection technique for RandomForest. **Novelty:** In this research, the AUROC score is slightly increased to 0.9974, which has not been achieved yet. Instead of 40, the number of features chosen is 20. This research also provides a comparison of different feature selection techniques like tree-based feature selection, information gain, Bhattacharya, and RFE. It also analyses their impact on the performance of models, which has not been done yet with the same set of selection techniques.

Keywords: Sepsis; Machine Learning; Feature selection; Early prediction; Predictive Analytics

1 Introduction

Sepsis is a life-threatening condition that can lead to failure of one or more organs [1]. Sepsis is the main cause of mortality in ICUs, and in the last decade, its occurrence has folded twice [2]. Sepsis is the costliest health care problem in the USA. Sepsis patients must be identified as soon as possible in order to limit mortality and morbidity [3].

To identify sepsis, a quick Sequential Organ Failure Assessment (qSOFA) score is useful. qSOFA Criteria are as follows: Altered mental status, a respiratory rate greater than 22 per minute, and SBP less than 100 mm of Mercury. qSOFA is positive if any two criteria are positive. When the qSOFA score is combined with organ dysfunction and infection, the chances of Sepsis are increased [2].

Various prediction systems are used in hospitals to predict sepsis. When rule-based systems (RBS) are compared with machine learning-based prediction methodology, the latter can reduce mortality by quite a large variance [4]. The number of false positives is higher in RBS, leading to more false alarms [5]. Automated Sepsis alarms based on electronic health records do not seem to reduce mortality or length of stay in hospitals [6]. Sensitivity, AUROC, and specificity for Epic Sepsis Prediction systems are 0.33, 0.63, and 0.83, respectively [7]. ML models seem to outperform for predicting Sepsis [8]. So, a machine learning approach is preferred.

A Literature Review (LR) has been conducted to find the different machine learning models predicting sepsis. It also caters to the feature selection methods used. For the current LR, a total of 35 research papers/review papers/articles, etc. were identified. Some articles were excluded since these belonged to sepsis detection. Also, a few research papers were excluded since these did not include a Machine learning approach and used expert systems. Similarly, few studies were rejected since these could not specify the dataset being used. A summary of chosen papers is shown in Table 1. Few papers used MIMIC II and/or MIMIC III datasets; however, some studies used data from some private hospitals. A list of different machine learning models that were used is as follows: Logistic Regression (LR), Support Vector Machine (SVM), Naïve Bayes, Gradient Boost, etc. All of these studies have confirmed their own AUROC score. The input features used by these sets were also different. Few research studies used notes entered by nurses/doctors. Few studies used only Vitals for prediction. Few research studies made use of blood tests done in hospitals. Studies undergone are as follows: [9].

The maximum AUROC for sepsis prediction in these studies is 0.98 [10]. Further, none of the studies have shown the impact of different feature selection techniques on different machine learning algorithms. For a few studies, feature selection techniques are not specified [11]. Few research works have not used feature selection at all [11–13]. Few research works have selected features manually, and then the selected features were fed to different machine learning algorithms to check the outcome [9,10]. Few research works have used different types of feature selection techniques to select the top features and then used the same in different models [14–16]. One of the research papers [17] investigated the impact of using different feature selection techniques on different ML algorithms, however, the feature selection techniques are fewer in number (only 3).

Table 1. Comparison Table

Reference	Dataset used	ML Method/Model	Features Used	Selection Methods used	AUROC	Sensitivity, Specificity
[9]	MIMIC-IV	Decision Tree, Gradient Boosting, XGBoost, LightGBM, MLP, SVM	29	Manual Feature Selection	0.94	0.8372, 0.8631
[11]	German tertiary care centre	Boosted Random Forest	7	None	0.872	-
[10]	FHC and Cardiology Challenge	XGBoost, SVM, RandomForest, ANN, Naïve Bayes	-	Manual Feature Selection	0.89 and 0.98	0.7790, 0.8442
[14]	MIMICIII	Gaussian Naïve Bayes (NB), Decision Tree, Random Forest, Logistic Regression (LR), KNN algorithm, and XGBoost	44	RandomForest model	0.918	0.9999, 0.9987
[17]	Multiple	Ensemble Techniques: ANN, RandomForest, SVM, Adaboost, LR, GB, KNN etc.	73, 47, 76 etc.	Info gain, Gini, Relief	0.918	0.852

Continued on next page

Table 1 continued

[12]	Singapore government-based hospital	SERA Algorithm	physicians' clinical notes with structured EMR data	None	0.94	0.89, 0.85
[15]	First Affiliated Hospital of Zhengzhou University	RandomForest	20	Gini Importance	0.91	0.87, 0.89
[16]	MIMIC II, MIMIC III, Emory University Hospital	SVM, decision trees, naive Gaussian Bayes and ensemble learners The first model evaluated patient features extracted during the first 56 hours using the AdaBoost ensemble technique along with a decision tree. Next, the discriminant subspace-based ensemble technique was used	24	PCA	0.489	-
[13]	Northern Lincolnshire, York Hospital and Goole Hospital	Logistic Regression Model	Vital signs and blood test results.	None	0.78	-

2 Methodology

Sepsis Criteria for Patients:

This section elaborates on the criteria of Sepsis for patients. Here is the list of time points for each patient:

1. Time of Sepsis suspicion: ($t_{suspicion}$).

It refers to the time of infection suspicion based on antibiotics.

2. Time of SOFA score: t_{SOFA}.

It refers to the time when there is a deterioration of 2 points in SOFA score within a day, which states the occurrence of organ damage.

3. Time of onset sepsis: t_{sepsis}. It refers to the sepsis onset time.

If $_{suspicion} - 24 \leq t_{SOFA} \leq t_{suspicion}$ +12, then t_{sepsis} = min ($t_{suspicion}$, t_{SOFA}).

For sepsis patients, SepsisLabel is 1 if t $\geq$ tsepsis—6 and 0 if t < tsepsis—6 For non-sepsis patients, SepsisLabel is 0.

2.1 Dataset Description

The dataset consists of patients' data from two hospitals' ICUs[18]. There were 2 training sets, 'A' and 'B' which comprised of multiple psv files. Each file has data from one patient only. The total data consists of 20336 number of patients and the total count was 790215 rows and 40 columns. The different categories of data is as follows: Vital Demographic and Laboratory data. Laboratory data contained information like Calcium level, Glucose level, Hgb (hemoglobin), etc. Demographic data included data like age, gender, length of stay at ICU, etc. Vitals data included data like Pulse oximetry, heart rate, etc. The output feature consists of data for Sepsis, if the value of this feature was 1, it indicated that the patients were predicted for Sepsis 6 hours in advance, otherwise, it was set as 0.

There was a huge missing data, which was analyzed further and filled in accordingly. Also, the highly imbalanced data was rectified. Details of the implementation are provided here.

2.2 Missing Data Analysis and Handling

Missingno python library was used to analyze missing data.

- The column EtCO2 did not have any value-filled, so it was removed from the dataset.
- The following features have minimal/no missing values: ICULOS, Gender, Age, SepsisLabel (Output).
- Few features had a higher number of missing values which include features like AST, Troponinl, EtCO2, etc.

For handling missing data, the following techniques were applied:

- Removed the columns which contain mostly missing data.
- The forward fill technique was applied per patient to compute a few missing values.
- Next, the backward fill technique was applied per patient to compute a few missing values.
- These techniques were applied for values applicable to a single patient, so as to avoid merging of data across multiple patients.

2.3 Imbalanced Data Handling

This case contained highly imbalanced data. Only 2% of patients were with Sepsis, 98% were non-Sepsis patients. The SmoteTomek technique was applied in this research paper for handling this case. SmoteTomek is a mixture of over-sampling and under-sampling techniques. SMOTE increases the minority samples, which in this case are patients with Sepsis. Tomek is an under-sampling technique that reduces the majority of samples, which in this case are the non-Sepsis patients. Subsequently, the NearMiss technique was also applied. NearMiss is an under-sampling technique that reduces the majority of samples so that the samples are in the specified ratio.

2.4 Feature Selection

Different feature selection techniques were applied to check the impact of these techniques on different machine learning models. Different filter, embedded, and wrapper techniques were applied. Filter methods use statistical measures to rank the importance of features. Some of the examples of filter methods are Correlation-based feature selection (CFS), Chi-Squared feature selection, and Mutual Information based feature selection. Among filter methods, Information Gain, and feature selection based on Bhattacharya distance were applied to models. In the information gain selection method, the information gain is calculated for each feature, which is then evaluated further with respect to the target variable. Figure 1 illustrates the importance of different features using the information gain technique.

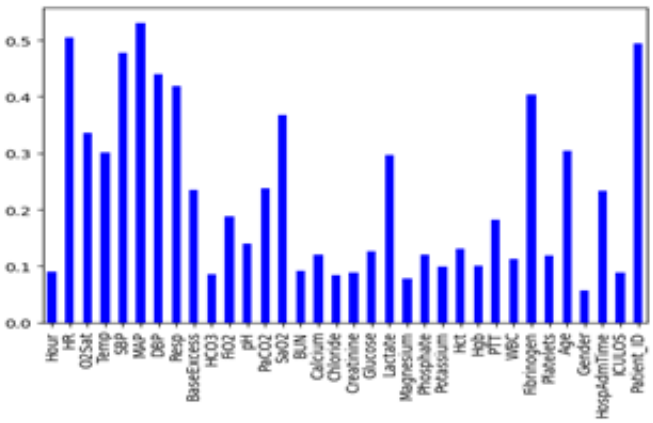

Fig 1. Feature importance – Information Gain

As seen in Figure 1, a few features like HR, SBP, MAP, Fibinogen, etc. had higher feature importance. Less important features were Gender, Magnesium, HCO3, etc.

Bhattacharya distance measures the similarity between probability distributions. Figure 2 shows the importance of different features based on Bhattacharya distance:

Figure 2 indicates that features like ICULOS, Hospital Admission Time, and Hour were most important. However, features like Age, Gender, MAP, etc. were not important for the prediction process.

Embedded methods: these methods combine feature selection with the model training process. Some examples of embedded methods are Lasso Regression, Ridge Regression, Elastic Net: Decision Trees, and, Random Forest. Among these algorithms, Lasso Regression and tree-based feature selection were implemented to select the top 20 features out of 40.

As observed in Figure 3, features like ICULOS, HospAdmTime, Hour, Temp, etc. were of higher importance than features like Platelets, Magnesium, PaCO2, etc.

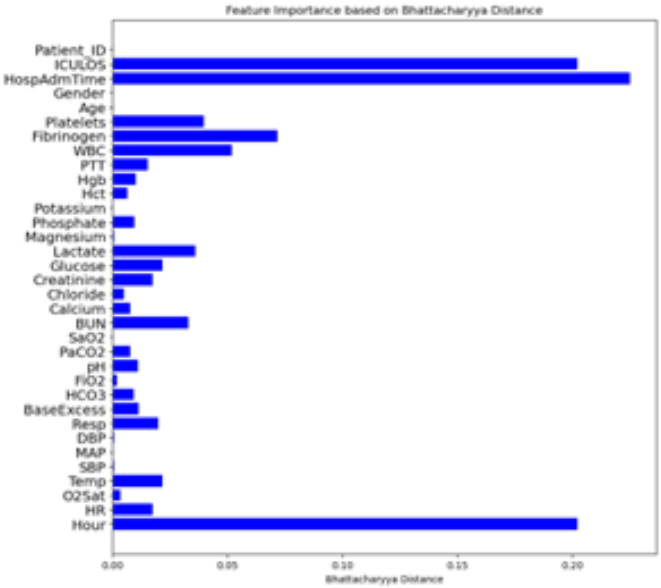

Fig 2. Feature importance based on Bhattacharya Distance

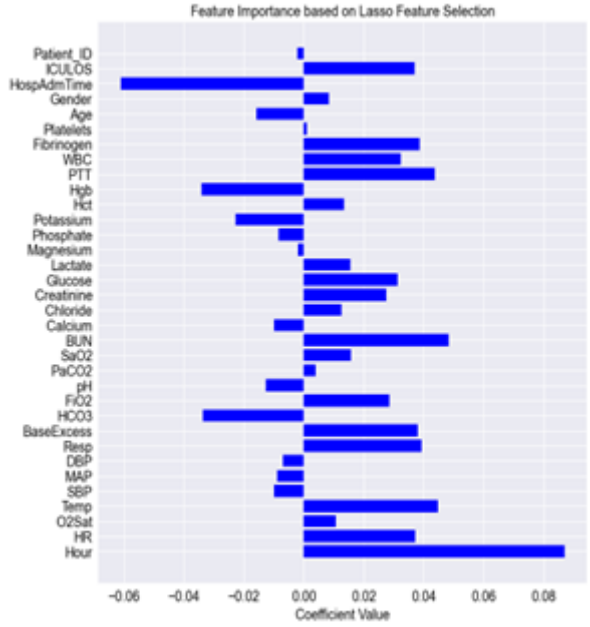

Fig 3. Feature importance based on Lasso Regression

Figure 4 shows that features like SaO2, Hct, Calcium, and Hgb are of least importance, however, features like Creatinine, Temp, BUN, etc. are of more importance.

Wrapper methods: these methods use the machine learning model's performance as a criterion for selecting the best subset of features. Some of the examples of wrapper methods most commonly used in various research papers are: Forward selection and backward elimination. Recursive feature elimination was used. This model keeps on deleting the weakest feature from the dataset, and it keeps on iterating until we have the required number of selected features out of the total features in the dataset.

Figure 5 shows that features like Gender, MAP, SBP DBP, etc. are of more importance than features like Hour, HospAdmTime, Age, Platelets, etc.

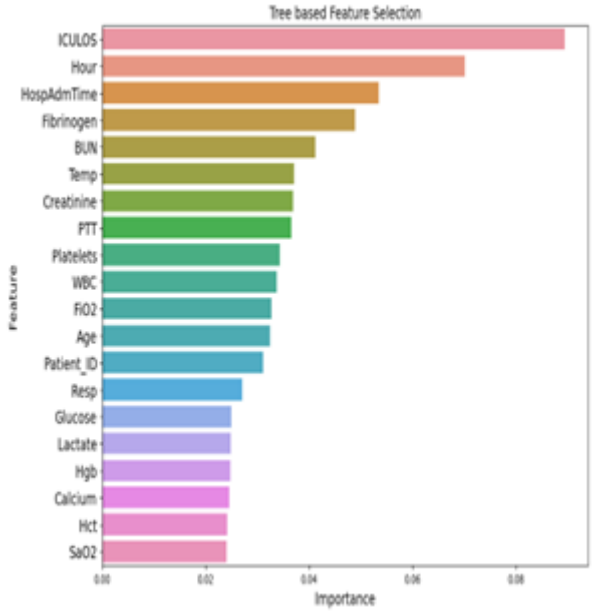

Fig 4. Tree-based feature importance

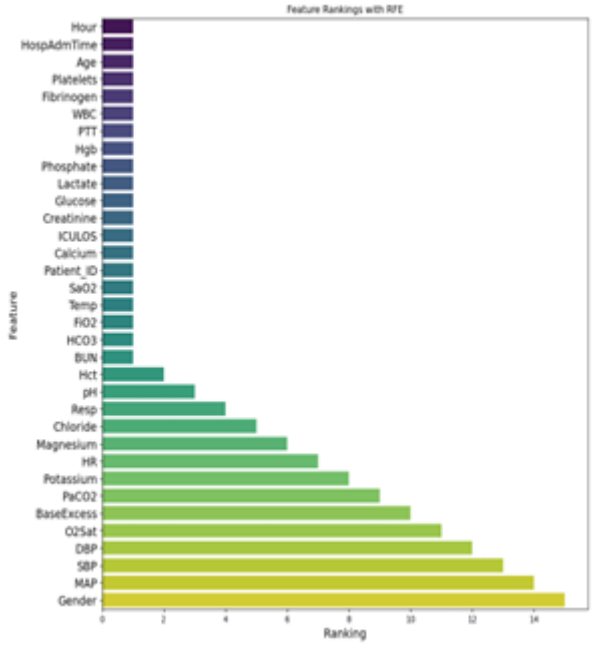

Fig 5. Feature importance based on Recursive Feature Elimination

2.5 Model Implementation and Evaluation

On the given dataset, a training sheet was prepared on a predefined ratio. The rest of the data was included for testing. So as to avoid overfitting, a stratified K-fold cross-validation technique was applied. Different hyperparameters were tuned for various models to obtain the best results. Based on earlier research [19], top 3 performing algorithms were used in this research paper, among the following models: Extreme Gradient Boosting (XGB), Gradient Boost (GB), AdaBoost (AB), Logistic Regression (LR), Easy Ensemble (EE), Stochastic Gradient Descent (SGD), etc.

3 Results and Discussion

Many feature selection techniques were applied to top-performing machine learning models that we got from earlier predictions. The following 3 algorithms were selected out of a total of 12 algorithms. These top 3 algorithms were Decision Tree, K Nearest Neighbors (KNN), and Random Forest. Different feature selection techniques were applied to these models. After applying these techniques, we again evaluated these models on the selected data. Evaluation of the algorithms was done on multiple measures, including AUROC, precision, recall (sensitivity), specificity, F-measure, and ROC curve.

3.1 Model Comparison of Performance

In the current research, 5 different feature selection techniques are applied to 3 ML models, making a total comparison of 15 models. As mentioned in Table 2, Sensitivity stands best for KNN with a tree-based selection technique. Random Forest with Recursive feature elimination topped when other performance parameters are considered like AUROC, Accuracy, etc.

Table 2. Models Performance Comparison with Selection Algorithms

S. No.	Model Name	Selection Algorithm	AUROC	Sensitivity	Specificity	Accuracy	F measure score
1	Random Forest	Information Gain	0.9890406	0.7394222	0.9971284	0.9915403	0.7912568
2	Random Forest	Bhattacharya	0.9970563	0.8981617	0.9976134	0.9954569	0.8955484
3	Random Forest	Tree based feature selection	0.9970683	0.9197549	0.9974841	0.9957986	0.9047072
4	Random Forest	Lasso	0.9972054	0.9025387	0.9975876	0.9955265	0.8974322
5	Random Forest	Recursive Feature Elimination	**0.9973955**	0.9171287	**0.9976264**	**0.9958809**	**0.9061554**
6	K Nearest Neighbors	Information Gain	0.9481682	0.8689816	0.9756946	0.9733807	0.5860474
7	K Nearest Neighbors	Bhattacharya	0.9523936	0.9034141	0.9698026	0.9683630	0.5532523
8	K Nearest Neighbors	Tree based feature selection	0.9751961	**0.9343449**	0.9858294	0.9847130	0.7260771
9	K Nearest Neighbors	Lasso	0.9719581	0.9217975	0.9856983	0.9843174	0.71754685
10	K Nearest Neighbors	Recursive Feature Elimination	0.9750535	0.9337613	0.98584235	0.9847130	0.7259528
11	Decision Tree	Information Gain	0.7954780	0.6311642	0.959816578	0.952690097	0.3665170
12	Decision Tree	Bhattacharya	0.8410625	0.7143274	0.967797641	0.962301399	0.4510779
13	Decision Tree	Tree based feature selection	0.9140079	0.8540998	0.973916024	0.971317932	0.5635891
14	Decision Tree	Lasso	0.8924385	0.8117887	0.973088167	0.969590554	0.5365477
15	Decision Tree	Recursive Feature Elimination	0.9082046	0.8427196	0.973689657	0.970849705	0.5562939

AUROC score is 0.9973955, which is the highest among all the literature present and has not been achieved so far.

When comparing the performance of models before [19] and after applying feature selection techniques, it is found that the performance degraded when information gain and Bhattacharya techniques were applied to decision trees and KNN models. However, performance improved when tree-based, lasso, and RFE techniques were applied to decision trees and KNN models.

Within these also, the tree-based selection technique showed the best performance. For the RandomForest model, results are a bit different; AUROC improved in all selection techniques, but sensitivity decreased when the information gain technique was applied. Rest techniques improved sensitivity. Specificity decreased slightly from 0.998 to 0.997 when feature selection techniques were applied to the RandomForest model.

3.2 ROC Curve Comparison

ROC (Receiver's Operating Characteristic Curve) curve plots the graph between False Positive Rate and True Positive Rate. ROC shows how a model performs at different thresholds. It is a probabilistic curve. Figure 6 illustrates the comparison of the ROC curve for all models with the defined feature selection techniques.

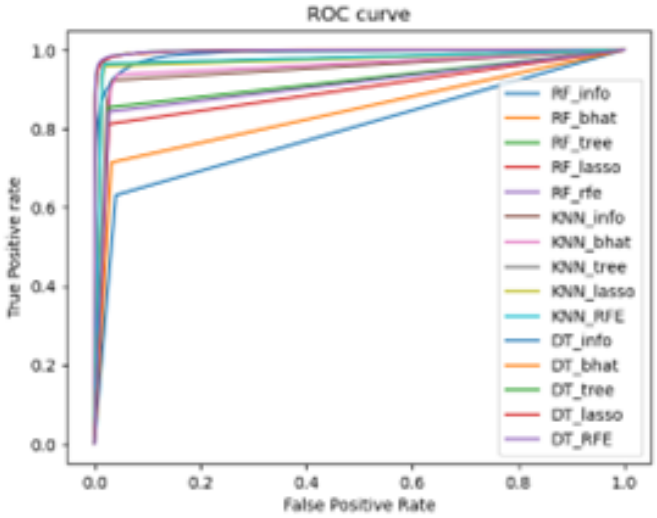

Fig 6. Performance comparison based on ROC

3.3 AUC Comparison

AUC is Area under ROC curve. It refers to the capability to clearly segregate positive and negative classes. The AUC score is directly proportional to the ML model performance.

The highest AUROC came for Random Forest with RFE selection technique, and the lowest came for Decision Tree with information gain and Bhattacharya selection techniques. The AUROC for the rest of the models seems to be quite comparable. The highest AUROC came out as 0.9974, which has not been achieved yet in the mentioned studies and research. After applying feature selection techniques, the AUROC decreased from 0.9965 to 0.9890 when feature selection using information gain was applied to the RandomForest model. However, the performance of other feature selection techniques has increased, for RandomForest, the AUROC got increased from 0.9965 to 0.9974 using RFE.

The highest AUROC achieved i.e. 0.9974, is based on 20 selected features. With the same count of features, the research [15] was able to get the AUROC of 0.91. Few research items took more features [9,16], however, the AUROC score is less than 0.9974. Other research studies took fewer features [11], however the AUROC score is less than 0.9974.

3.4 Precision Comparison

It is the ratio of correct predictions out of all actual positive items.

The precision of Random Forest is highest with a few selection methods like Bhattacharya, Lasso regression, and RFE. Among the selection algorithms, information gain seems to be lowest in our case when applied to the Decision Tree algorithm. In the case of KNN, the lowest precision seems to be of the Bhattacharya algorithm.

3.5 Accuracy Comparison

It signifies the total items that are predicted correctly out of all of the total items. The correct predictions are TN + TP. Total items are as follows: TP + FP + TN + FN. So, this is a ratio of these 2 items.

If we only see accuracy, then the performance of all selected models for the selected selection algorithms is comparable. The highest accuracy came out as 0.9959 for Random Forest with the RFE feature selection technique. After applying feature selection techniques, the accuracy decreased from 0.9940 to 0.9915 while feature selection using information gain was applied to the Random Forest model. However, the performance of other feature selection techniques has increased, like for Random Forest, whose accuracy got increased from 0.9940 to 0.9959 using RFE.

3.6 Error rate comparison

It specifies the ratio of false predictions (FN + FP) from total predictions, i.e., TP + FP + TN + FN.

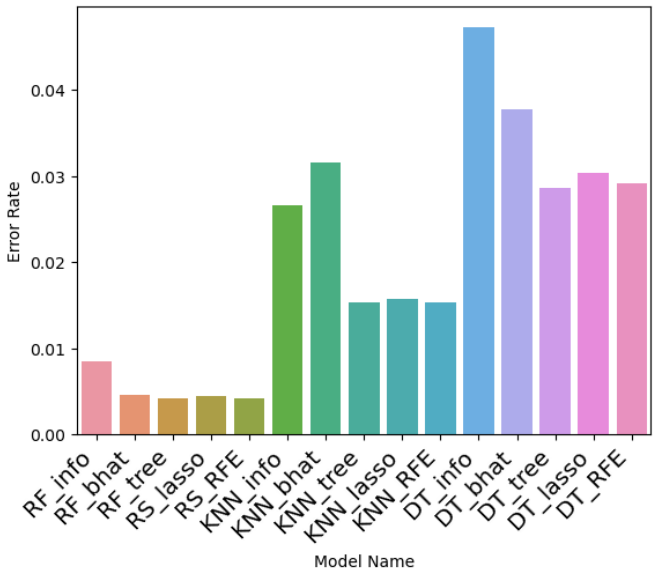

Fig 7. Performance comparison based on Error Rate

As seen in Figure 7, the highest error seems to be of DT with information gain and Bhattacharya selection techniques. The top-performing model in terms of error rate is the RandomForest model with RFE and tree-based selection techniques.

3.7 Sensitivity/Recall comparison

It signifies the ratio of positive predictions that are identified correctly, i.e., TP, out of actual positive items, which is a sum of TP and FN.

The highest recall came after applying the tree-based selection feature technique to the K Nearest Neighbor model. The lowest performance is shown by the Decision Tree with the information gain selection technique. The sensitivity of the KNN model stands best before and after applying feature selection techniques. The highest sensitivity came out as 0.9343 for KNN when the tree-based feature selection technique was applied. After applying feature selection techniques, sensitivity decreased from 0.9218 to 0.8690 when feature selection using information gain and Bhattacharya techniques were applied to the KNN model. However, for KNN, sensitivity increased from 0.9218 to 0.9338 and 0.9343 using RFE and the tree-based feature selection techniques, respectively. The sensitivity remained the same while lasso-based feature selection was applied to KNN (0.9218).

3.8 Overall comparison

A total comparison of all models with different selection algorithms has been provided in Figure 8. It clearly shows that the RandomForest model with RFE topped the matrix.

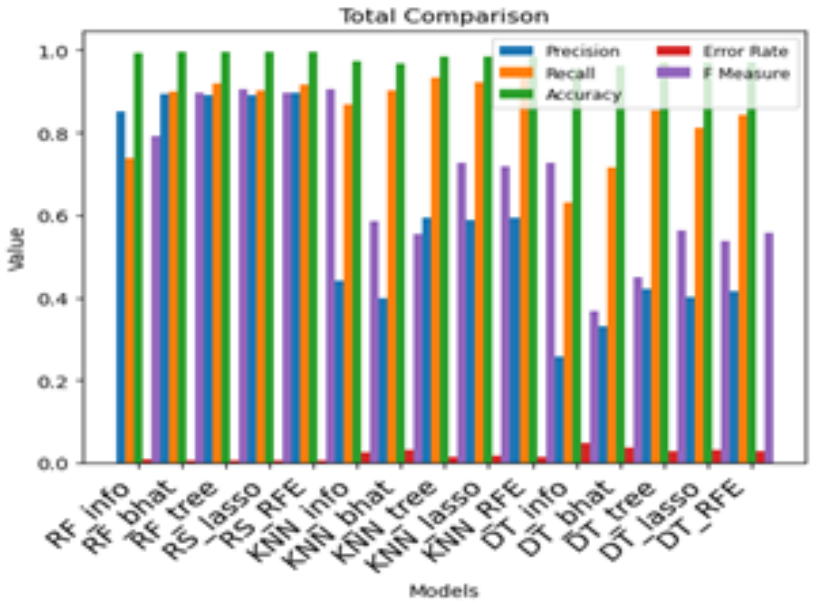

Fig 8. Model's Overall Comparison

Overall, AUROC, specificity, accuracy, and f-measure score stand best for the Random Forest model with RFE technique. Sensitivity stands best for the KNN model with a tree-based selection algorithm.

3.9 Performance of different selection techniques

Overall, the application of the RFE technique enhanced the performance of all models (KNN, DT, RandomForest) across all considered performance matrices like AUROC, Sensitivity, Specificity, F-measure score, etc. The information gain technique has slightly decreased the performance of all models (KNN, DT, RandomForest) across all considered performance matrices like AUROC, Sensitivity, Specificity, F-measure score, etc. Bhattacharya technique has increased AUROC, Sensitivity, and F-measure score, however, it has decreased specificity from 0.9981 to 0.9976 when applied to Random Forest. For KNN and DT, the performance decreased when the Bhattacharya technique was applied. So, the Bhattacharya technique is better suited for the Random Forest model as compared to KNN and DT models. The Lasso technique has improved the performance of the DT model across all considered performance parameters. When this technique was applied to the KNN model, the performance remained almost the same as it was without applying the technique. When this technique was applied to RandomForest, the performance of the model improved a bit except for specificity, where it decreased from 0.9981 to 0.9976. The performance of KNN and DT models has improved across all considered performance criteria after applying the tree-based selection technique. However, when this technique was applied to the Random Forest model, the performance decreased in terms of Specificity from 0.9981 to 0.9975 but increased across the rest of the performance criteria.

When comparing the sensitivity and specificity of the different research articles, sensitivity and specificity were not defined for a few studies[11,13,16]. When comparing our work with the studies, we have found that one of the research articles[14] has higher sensitivity and specificity at 0.9999 and 0.9987 respectively. Rest studies have lower values of sensitivity and specificity[9,10,12,15,17].

4 Conclusion

Sepsis is a serious disease that can quickly reach up to organ dysfunction. This study assessed different ML algorithms along with the feature selection impact on the output of the model, including AUROC, recall, precision, F-measure, accuracy, and error rate. The impact of selection techniques on different machine learning algorithms is identified. The AUROC score range lay between 0.7955 and 0.9974. The selection techniques like RFE, and tree-based selection techniques topped the score since the highest performance in terms of AUROC (0.9974), specificity (0.9976), accuracy (0.9959), and f-measure score (0.9062) is achieved when RFE is applied to the RandomForest model. Also, sensitivity (0.9343) stands best for KNN when the tree-based

feature selection technique is applied. Tree-based selection techniques work quite well on tree-based models like RandomForest and Decision Trees. In the case of RandomForest, AUROC increased from 0.9965 to 0.9971, and in the case of Decision Tree, AUROC increased from 0.8542 to 0.9140 when the tree-based selection technique was applied. The lowest performing model with respect to error is the Decision Tree model when the information gain feature selection technique is applied. The highest error appears in this case, i.e., 0.0473.

The number of features that have been considered in different selection models is 20, which is less than the total features, i.e., 40. Doctors can now take these features into consideration for predicting sepsis.

In the future, other selection methods will be included along with other machine learning models. Further, to avoid bias, data from Indian hospitals can be taken. The current dataset is not externally validated.

This research suggests that hospitals can use machine learning models along with feature selection to improve prediction performance. If Sepsis is predicted earlier, then medical expenses could be reduced along with a decrease in mortality and morbidity.

References

1) Srzić I. Sepsis definition: What's new in the Treatment Guidelines. *Acta clinica croatica*. 2022;61(1):67–67. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9536156.
2) Gul F, Arslantas MK, Cinel I, Kumar A. Changing definitions of sepsis. *Turkish Journal of Anesthesia and Reanimation*. 2017;45(3):129–167. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5512390/.
3) Gyawali B, Ramakrishna K, Dhamoon AS. Sepsis: The Evolution in Definition, Pathophysiology, and Management. *SAGE Open Med [Internet]*. 2019;7. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6429642/.
4) Zhang Z, Chen L, Xu P, Wang Q, Zhang J, Chen K, et al. Effectiveness of automated alerting system compared to usual care for the management of sepsis. *NPJ Digital Medicine [Internet]*. 2022;5(1):1–10. Available from: https://www.nature.com/articles/s41746-022-00650-5.
5) Joshi M, Mecklai K, Rozenblum R, Samal L. Implementation approaches and barriers for rule-based and machine learning-based sepsis risk prediction tools: a qualitative study. *JAMIA Open [Internet]*. 2022;5(2). Available from: http://dx.doi.org/10.1093/jamiaopen/ooac022.
6) Makam AN, Nguyen OK, Auerbach AD. Diagnostic accuracy and effectiveness of automated electronic sepsis alert systems: A systematic review. *Journal of Hospital Medicine [Internet]*. 2015;10(6):396–402. Available from: https://pubmed.ncbi.nlm.nih.gov/25758641/.
7) Habib AR, Lin AL, Grant RW. The epic sepsis model falls short-the importance of external validation. *JAMA Internal Medicine [Internet]*. 2021;181(8). Available from: https://pubmed.ncbi.nlm.nih.gov/34152360/.
8) Moor M, Rieck B, Horn M, Jutzeler CR, Borgwardt K. Early prediction of sepsis in the ICU using machine learning: A systematic review. *Frontiers in Medicine*. 2021;8(8). Available from: http://dx.doi.org/10.3389/fmed.2021.607952.
9) Gao J, Lu Y, Domingo IR, Alaei K, Pishgar M. Predicting Sepsis Mortality Using Machine Learning Methods. *medRxiv (Cold Spring Harbor Laboratory)*. 2024. Available from: http://medrxiv.org/content/early/2024/03/16/2024.03.14.24304184.abstract.
10) Wu Q, Ye F, Gu Q, Shao F, Long X, Zhan Z, et al. A Customised Down-sampling Machine Learning Approach for Sepsis Prediction. *International Journal of Medical Informatics*. 2024;184:105365–105370. Available from: https://academic.oup.com/clinchem/article/70/3/506/7618099.
11) Steinbach D, Ahrens PC, Schmidt M, Federbusch M, Heuft L, Lübbert C, et al. Applying Machine Learning to Blood Count Data Predicts Sepsis with ICU Admission. *Clinical chemistry*. 2024;70(3):506–521. Available from: https://pubmed.ncbi.nlm.nih.gov/38350181/.
12) Goh KH, Wang L, Yeow A, Poh H, Li K, Yeow J, et al. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nature Communications*. 2021;12(1):711–711. Available from: https://doi.org/10.1016/j.imu.2023.101236.
13) Faisal M, Scally A, Richardson D, Beatson K, Howes R, Speed K, et al. Development and external validation of an automated computer-aided risk score for predicting sepsis in emergency medical admissions using the patient's first electronically recorded vital signs and blood test results. *Critical Care Medicine [Internet]*. 2018;46(4):612–620. Available from: https://doi.org/10.3390/electronics11091507.
14) Gholamzadeh M, Abtahi H, Safdari R. Comparison of different machine learning algorithms to classify patients suspected of having sepsis infection in the intensive care unit. . *Informatics in Medicine Unlocked*. 2023;38:101236–101236. Available from: https://doi.org/10.1016/j.imu.2023.101236.
15) Wang D, Li J, Sun Y, Ding X, Zhang X, Liu S, et al. A machine learning model for accurate prediction of sepsis in ICU patients. *Front Public Health*. 2021;9. Available from: http://dx.doi.org/10.3389/fpubh.2021.754348.
16) Abromavičius V, Plonis D, Tarasevičius D, Serackis A. Two-stage monitoring of patients in Intensive Care Unit for sepsis prediction using non-overfitted machine learning models. . *Electronics (Basel)*. 2020;9:1133–1133. Available from: https://www.mdpi.com/2079-9292/9/7/1133.
17) Camacho-Cogollo JE, Bonet I, Gil B, Iadanza E. Machine Learning Models for Early Prediction of Sepsis on Large Healthcare Datasets. . *Electronics*. 2022;11:1507–1507. Available from: https://doi.org/10.3390/electronics11091507.
18) Rhee C, Jones TM, Hamad Y, Pande A, Varon J, O'brien C, et al. Prevalence, Underlying Causes, and Preventability of Sepsis-Associated Mortality in US Acute Care Hospitals. *JAMA Network Open*. 2019;2(2). Available from: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2724768.
19) Vandana, Hari PB. Comparison of machine learning models for predictive analytics of sepsis in the emergency medical admissions. *5th International Conference on Inventive Research in Computing Applications (ICIRCA) IEEE; 2023*. 2023;p. 38–45. Available from: https://ieeexplore.ieee.org/document/10220741.