Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 20 No. 4 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v20i04.45413

PAPER

Enhancing Classification Performance through FeatureBoostThyro: A Comparative Study of Machine Learning Algorithms and Feature Selection

Deepali Bhende[1](✉), Gopal Sakarkar[1,2], Punam Khandar[3], Satyajit Uparkar[3], Arvind Bhave[4]

[1]G. H. Raisoni University, Saikheda, Madhya Pradesh, India

[2]Dr. Vishwanath Karad, MIT World Peace University, Pune, Maharashtra, India

[3]Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, India

[4]Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur, Maharashtra, India

deepali.bhende.phdcs@ghru.edu.in

ABSTRACT

Early-stage prediction of a disease is an important and challenging task. The application of machine learning techniques is playing an important role in this era. Thyroid is one of the chronic endocrine diseases, and approximately 42 million people in India are affected by this disease. This paper presents a comprehensive investigation into the enhancement of classification performance through the novel 'FeatureBoostThyro' (FBT) model. The study evaluates various machine learning algorithms, including stochastic gradient descent (SGD), K nearest neighbor (KNN), logistic regression (LR), naive bayes (NB), and support vector machine (SVM), in conjunction with diverse feature selection methods. The research systematically explores the impact of feature selection techniques such as information gain, relief F, chi-square, gini index, forward selection, backward selection, recursive feature elimination, and LASSO on model performance across the chosen algorithms. The analysis reveals notable variations in performance metrics, including accuracy, precision, recall, and F1-score, providing valuable insights into the interplay between algorithm and feature selection. One main contribution of this research is the introduction of the FBT model, which consistently outperforms other models across various feature selection methods, making it a promising tool for addressing complex classification tasks. The findings contribute to a broader understanding of model selection and optimization in machine learning applications. The proposed model undergoes evaluation using two distinct datasets: the primary dataset acquired from Lata Mangeshkar Hospital in Nagpur and the secondary dataset obtained from the UCI dataset.

KEYWORDS

machine learning, thyroid disorder, feature selection, logistic regression, gradient descent, information gain, recursive feature elimination

Bhende, D., Sakarkar, G., Khandar, P., Uparkar, S., Bhave, A. (2024). Enhancing Classification Performance through FeatureBoostThyro: A Comparative Study of Machine Learning Algorithms and Feature Selection. *International Journal of Online and Biomedical Engineering (iJOE)*, 20(4), pp. 29–42. https://doi.org/10.3991/ijoe.v20i04.45413

Article submitted 2023-09-30. Revision uploaded 2023-12-11. Final acceptance 2023-12-11.

1 INTRODUCTION

There are several chronic diseases, and treating these chronic diseases is a difficult challenge for doctors [1]. One of the chronic disorders is thyroid disease. The thyroid is located on the front side of the neck, and its malfunction results in this disease. The trachea is surrounded by endocrine gland. It has the shape of a butterfly. The thyroid gland produces the hormone that regulates many important bodily functions. Thyroid illness occurs when the thyroid gland fails to produce the appropriate amount of thyroid hormones. These hormones' function is to keep the body running smoothly. When too much thyroid hormone is secreted, the body expends energy too quickly. This disease is known as hyperthyroidism, and it causes fatigue as a result of burning energy too quickly. It also creates a quick heartbeat, which induces weight loss without effort, as well as an anxious feeling. When the thyroid gland secretes insufficient thyroid hormone, the disease is known as hypothyroidism. When the body produces insufficient thyroid hormone, one may experience fatigue. It may induce weight gain and possibly the inability to tolerate cold temperatures [2] to keep the body working regularly.

Hypothyroidism and hyperthyroidism are the two most common thyroid disorders. Both disorders affect the thyroid gland's function. Thyroiditis, Hashimoto's thyroiditis, postpartum thyroiditis, and other conditions can induce hypothyroidism. Grave's disease, nodules, thyroiditis, and other diseases can induce hyperthyroidism. Proper treatment of thyroid disease always reduces the high risk of death. To enhance diagnostic accuracy, machine learning techniques are employed. The study findings encompass a nuanced understanding of algorithm and feature selection interactions, culminating in the introduction of the innovative FBT model. This model's potential for real-world deployment and the comprehensive evaluation of methods make this study a valuable contribution to the field of machine learning and classification tasks. The model tested against both the primary and secondary datasets. The primary dataset was collected from Lata Mangeshkar Hospital in Nagpur, while the secondary thyroid data set collected from the UCI repository.

The paper is structured into various sections. In Section 2, the paper conducts a literature survey, where the primary focus is on studying alternative feature selection strategies and analyzing their influence on the performance of machine learning models in terms of accuracy. Moving on to Section 3, the proposed architecture is presented, offering comprehensive insights into the machine learning model used and the specific features of the thyroid dataset under investigation. Section 4, titled result analysis, is where the findings are presented and the outcomes of this study are interpreted. Finally, the paper concludes in the conclusion section, summarizing the key takeaways and implications of this study.

2 LITERATURE SURVEY

S. Nandinidevi et al. [3] discussed the importance of the random forest method as a feature selection method. In the study, the thyroid dataset was used, which is available at the UCI repository. The dataset consists of a total of 21 features, from which important features are extracted. The performance of the K-nearest neighbor machine learning algorithm was evaluated against a total of 21 features, and the accuracy recorded is 99.8%. After the implementation of the feature selection method, 100% accuracy is achieved.

Avijit Kumar Chaudhari et al., in their paper [4], mentioned that there is a lack of a singular data mining algorithm that consistently provides accurate results for health-care datasets. The study considers the heart disease dataset of a total of

155 patients from the UCI data repository. Different machine learning techniques used under consideration were random forest (RF), decision tree (DT), naïve bayes (NB), support vector machine (SVM), extra tree, gradient boosting, and logistic regression (LR). A comparative study has been carried out, which consists of a subset of features with 4, 6, 8, and 13 features against the different feature selection methods. The methods utilized in this study include one R, gain ratio, relief F, and Information gain. The study's conclusion emphasizes that the performance of a data mining technique is contingent upon the specific characteristics of the dataset being analyzed.

The study by Karna Vishnu Vardhana Reddy et al. [5] involved a comparison of the performance of ten machine learning classifiers along with three feature selection methods, namely chi-squared, relief F, and correlation-based method. For disease risk prediction, the dataset was collected from the Cleveland Heart dataset. The classification process was conducted using cross-validation, and multiple classifiers were utilized, including NB (naive bayes), LR (logistic regression), SMO (support vector machines with optimization), AdaBoost (adaptive boosting), bagging, and RF (random forest). Among all classifiers, the SMO classifier gave better performance with the chi-squared approach, and the NB method outperformed the correlation-based feature selection method.

To assess the performance of the proposed framework, Kapil Juneja et al. [6] used DT, decision table, RF, random tree method, NB, multilevel perceptron, and RBF networks to analyze the performance of the extreme learning machine composite fuzzy rule-based method. A total of five experiments were conducted using different FS methods, such as relief F, chi-square, gain ratio, and information gain. Two datasets, namely thyroid L7 and hypothyroid, were utilized in these experiments. Each of them consists of 3772 and 1972 instances, respectively. With the thyroid dataset, the proposed model achieved a maximum accuracy of 95.9%. J48, DT, and RF also achieved 99% accuracy. With the thyroid L7 dataset, among all methods, decision table, decision tree, random tree, and the random forest classifier achieved significant accuracies of 95.9%, 89.38%, 95.15%, and 91.46% on the respective datasets. Additionally, the proposed method demonstrated even higher accuracy, with a score of 96.25%.

R. Vartharajan et al. [7] proposed an enhanced method that includes a kernel-based SVM method combined with (LDA). The performance of this method was compared with LDA combined with (MLP), LDA combined with SVM, and (PCA) combined with SVM. The results showed that LDA with MLP achieved an accuracy of 84% and 79% on the respective datasets. PCA with SVM achieved an accuracy of 93% and 89%, respectively. However, LDA with enhanced SVM gives the best results as compared to other methods. The accuracy obtained was 96%, and the sensitivity and specificity values were 94%.

Dhyan Chandra Yadav et al. [8] designed a prediction model for heart disease using a dataset that is available at the UCI repository. The dataset comprises 1025 instances and a total of 14 attributes. The study focused on three tree-based classification techniques, namely M5P, random forest, and random tree, along with feature selection methods including Pearson correlation, Lasso regularisation and recursive feature elimination. There are three experimental setups. In the first setup, all ML algorithms are evaluated with the Pearson correlation method. In the second setup, they applied recursive feature elimination, and in the third setup, they used the method with different classifiers. The best results, i.e., 99% accuracy, are obtained by with the RF ensemble method.

Priyanka Sonar et al. [9] used four distinct machine learning algorithms to predict diabetic risk in patients, including SVM, NB, neural networks, and DT. Performance was evaluated using various criteria such as recall, precision, accuracy, support, and f1 score. When compared to DT and NB, the model performs better using SVM and ANN algorithms.

K Shankar et al. [10] developed one that incorporates a kernel-based classifier and optimal feature process. The feature selection method was employed to enhance the model performance. Multikernel SVM classifier achieved 97.49%, 99.05%, and 94.5% accuracy, sensitivity, and specificity, respectively.

In this literature review, various machine learning algorithms and their performance with different feature selection methods are examined. The results indicate significant improvements in accuracy after applying feature selection. These findings are summarized in Table 1, which underscores the crucial role of feature selection methods in improving the performance of machine learning algorithms, as evidenced by the improvements in accuracy (*after) compared to the initial results (*before).

Table 1. Report on literature review indicating use of FS Method, ML technique, and prediction accuracies

Ref No	Authors	ML Algorithms	Feature Selection Methods	Accuracy	
[3]	S. Nandini Devi et al. (2021)	KNN	Random Forest	99.66% (*before) 100% (*after)	
[4]	Avijit Kumar (2021)	NB, SVM, LR, DT, RF, GDB Extra Tree	Information Gain, Relief F, Gain Ratio, One R	For IG: 87.10%, 90.32%, 74.19%, 87.10%, 77.42%, 80.65%	
[5]	Karna Vishnu Vardhana Reddy (2021)	NB LR SMO KNN Adaboost Bagging RF	Correlation based FSM	Before 83.82% 84.81% 85.14% 76.89% 82.83% 80.85% 81.84%	After 84.15% 83.16% 83.82% 78.87% 83.82% 81.18% 79.53%
[6]	Kapil Juneja (2021)	Extreme LM Composite FR based method Using DT, DT, RF, RTM	Information Gain, Gain Ratio, Chi-square, Relief F	95.9%, 89.38%, 95.15%, 91.46% resp. EML: 96.25%	
[7]	R. Varathrajan (2021)	Enhanced SVM	Linear Discriminant Analysis (LDA)	96% (after)	
[8]	Dhyan Chandra Yadav et al. (2020)	RF	Pearson Correlation FSM, Recursive FSM	99.9% 94.12%	
[9]	Priyanka Duggal (2020)	RF NB SVM	Univariate Selection, Recursive Feature Elimination, Tree based FS	With RFE: 78.21%, 74.37%, 92.92% resp.	
[10]	K. Shankar (2020)	Multikernel SVM	Gray Wolf Optimization	94.55% (before) 99.02% (after)	

Notes: *before: before applying feature selection method; *after: after applying feature selection method.

3 PROPOSED ARCHITECTURE

The proposed system architecture assesses a variety of machine learning techniques, including stochastic gradient descent (SGD), K nearest neighbor (KNN), support vector machine (SVM), naive bayes (NB), and logistic regression (LR). This evaluation encompasses a range of performance metrics, including F1-score, accuracy, precision, and recall. The study broadens its scope by incorporating diverse feature selection methods, such as information gain, Gini index, chi-square, F, forward selection, backward selection, recursive feature elimination, and Lasso. The research thoroughly investigates the impact of applying these distinct feature selection techniques.

To further enhance the overall system performance, an Adaboost ensemble model is introduced in combination with the search optimization algorithm. This fusion aims to significantly bolster the system's performance. This innovative model is denoted as 'FeatureBoostThyro' (FBT). The experiments are executed meticulously to assess the performance of the newly introduced FBT when employed with various machine learning models. This evaluation employs a comprehensive set of performance evaluation metrics.

Data pre-processing serves as a pivotal initial stage within the data mining process, entailing the conversion of raw data into a format amenable to subsequent analysis. This step encompasses the management of missing or inconsistent values, ensuring that irregularities do not unduly impact the results. Additionally, the dataset undergoes partitioning through a splitting operation, resulting in separate training and testing datasets. The model is then trained using the training dataset, and the outcomes produced by the trained model are contrasted with the anticipated values present in the test dataset. Figure 1 offers a visual depiction of the overall framework structure.

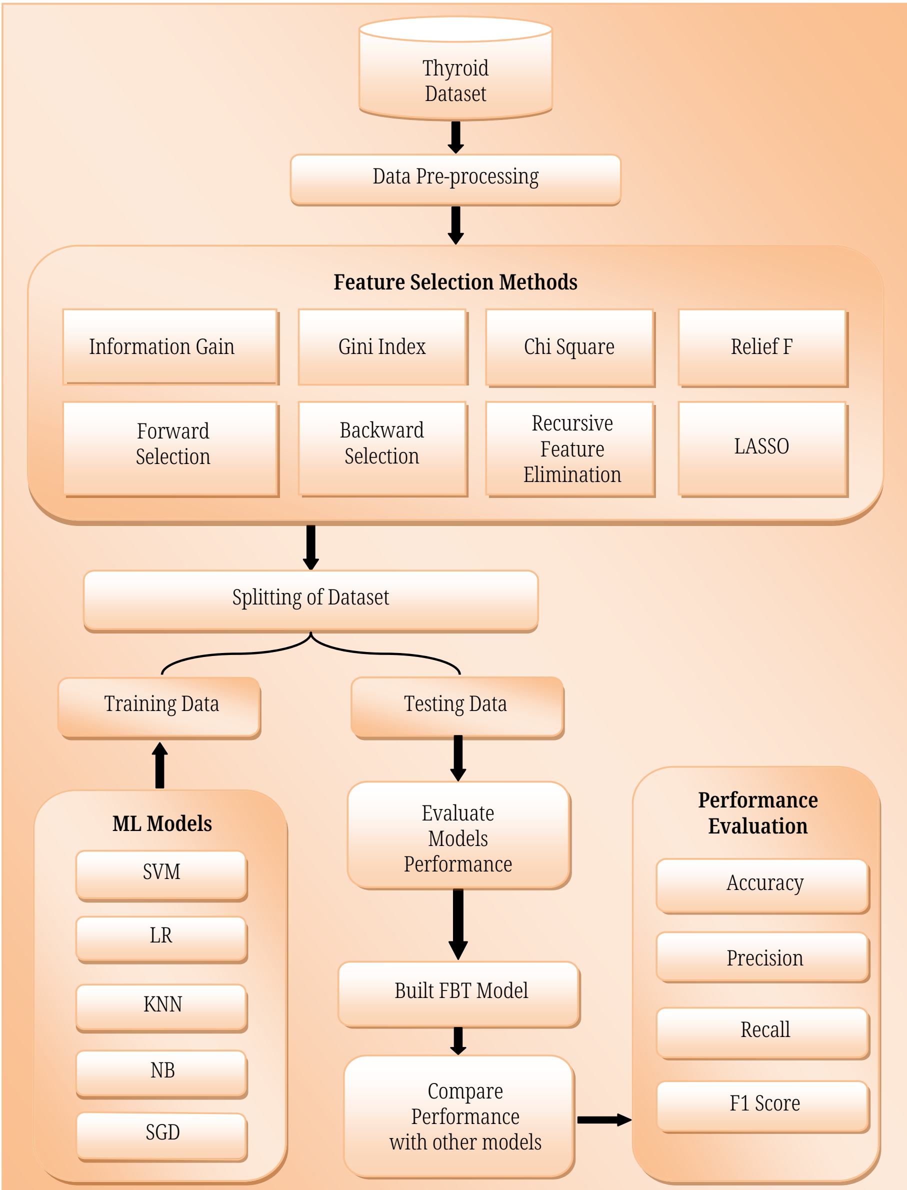

Fig. 1. Proposed framework system architecture

3.1 Data source

The secondary dataset for thyroid disease is downloaded from the UCI repository. There are a total of 3773 records and 29 attributes present in the dataset [11]. Table 2 shows the details of the attributes of the dataset. The thyroid disorder prediction model is based on the dataset.

Table 2. Attributes of thyroid dataset

Thyroid Dataset Features	
1. age	**16.** psych
2. sex	**17.** TSH measured
3. on thyroxine	**18.** TSH
4. query on thyroxine	**19.** T3 measured
5. on antithyroid medication	**20.** T3
6. sick	**21.** TT4 measured
7. pregnant	**22.** TT4
8. thyroid surgery	**23.** T4U measured
9. I131 treatment	**24.** T4U
10. query hypothyroid	**25.** FIT measured
11. query hyperthyroid	**26.** FIT
12. lithium	**27.** TBG measured
13. goitre	**28.** TBG
14. tumor	**29.** referral source
15. hypopituitary	**30.** binary class

The primary data set was collected from Lata Mangeshkar Hospital in Nagpur. It comprises 225 patient records encompassing eight features: age, sex, TSH, T3, T4, free T3, free T4, and result.

3.2 Machine learning algorithms

Many popular data mining algorithms are used for classification problems as described below.

Based on Bayes' theorem, NB is a simple and probabilistic machine-learning technique. Given the class label, it believes that features are independent. By multiplying the conditional probabilities of each feature value given that class, it determines the likelihood of a class label for a particular instance. NB calculates the likelihood of a class label, C, given the features, F, assuming feature independence:

$$P(C|F) = P(C) \text{ X } P(F1|C) \text{ X } P(F2|C) * \ldots * P(Fn|C)$$

Where:

P(C|F) is the conditional probability.

K nearest neighbor classifies a new instance by taking into account its K-nearest sample values from the training data. The neighbours are identified using a distance measure such as Euclidean distance. Majority vote (for classification) or average (for regression) among the K neighbors determines the projected class or value [12].

KNN classifies a new instance by considering its K-nearest samples, N, based on a distance metric, D (e.g., Euclidean distance). The predicted class, CP, is determined by majority voting (MV) for classification or averaging (AVG) for regression among the K neighbors:

$$N = \{\text{K-nearest samples}\}$$
$$CP = MV(N) \text{ or } AVG(N)$$

Where:
MV is majority voting.
AVG is averaging.
N is the set of nearest neighbors.
CP is the predicted class.

Using the logistic function, the logistic regression technique models the relationship between the independent and dependent variables. LR computes the likelihood that a given instance belongs to a specific class and then applies a decision threshold to generate predictions. Using the logistic function σ, LR models the likelihood, P, that a given instance belongs to a specific class, C.

$$P(C) = \sigma(W * X + b)$$

Where:
P(C) is the probability.
σ is the logistic function.

Support vector machines (SVM) is a powerful supervised machine learning approach for classification and regression applications [12]. It identifies the best hyperplane to maximize the separation between several data points. SVM aims to find the optimal hyperplane, H, that maximizes the margin, M, between data points, X, using a decision boundary function, f.

$$H: f(X) = 0$$
$$M = \max(\text{Margin})$$

Where:
H is the hyperplane
M is the margin

The optimization algorithm SGD is frequently used in machine learning to train models. It computes the gradient and changes the model parameters iteratively by selecting a sample subset of training samples at random. SGD is used to train machine learning models by iteratively updating model parameters, θ, based on the gradient, $\nabla\theta$, computed from a random subset of training samples, S.

$$\theta = \theta - \eta * \nabla\theta(\text{Loss})$$

Where:
θ is the model parameters
η is the learning rate
$\nabla\theta$ is the gradient
S is the subset of training samples
L is the loss function

4 RESULT ANALYSIS

Confusion matrices are utilized as performance indicators for numerous classification applications. As true negative (TN), true positive (TP), false positive (FP), and false negative (FN), it splits the values into four categories. There are four possible ways that the actual and predicted numbers of samples can be combined.

Different evaluation metrics are described below:

Accuracy: Accuracy is a metric used to estimate the proportion of correctly classified values. This approach entails dividing the entire number of instances (TP + TN + FP + FN) by the sum of TP and true negatives.

Precision: Precision is a metric that measures how accurate a model is at classifying positive values. It is calculated by dividing the total number of TP by the total number of false positives.

Recall: Recall, also known as the true positive rate, assesses the model's ability to properly forecast positive values. It is calculated by dividing the total number of TP by the total number of actual positive values (APV).

F1-Score: The F1-Score is used when it is necessary to find a balance between precision and recall. It provides a statistical metric that balances these two opposing abilities. The F1-Score formula involves multiplying recall and accuracy by two and dividing the result by the sum of recall and precision [(2 * Recall * accuracy) / (Re-call + Precision)].

The results of the performance of the ML techniques under examination are displayed in the following Table 3.

4.1 Experimental results for filter selection methods using UCI data set

Table 3 summarizes the results of the performance of different ML models and the proposed FBT model.

Table 3. Comparison of ML and FBT models with different filter feature selection methods

Feature Selection Method	Model Name	Accuracy	Precision	Recall	F1 Score
Information Gain	LR	96.37	96.28	96.37	95.68
	SVM	96.21	95.81	96.21	95.86
	KNN	94.15	88.65	94.15	91.32
	NB	94.15	88.64	94.15	91.32
	SGD	94.15	88.65	94.15	91.32
	FBT	**98.10**	**98.04**	**98.10**	**98.00**
Gini index	LR	94.15	88.65	94.15	91.32
	SVM	95.10	94.59	95.10	94.78
	KNN	94.31	92.82	94.31	91.97
	NB	29.23	92.14	29.23	38.81
	SGD	93.36	88.61	93.36	90.92
	FBT	**98.10**	**98.04**	**98.10**	**98.00**
Chi-square	LR	96.37	96.28	96.37	95.68
	SVM	96.52	96.21	96.52	96.14
	KNN	94.15	88.65	94.15	91.32
	NB	94.31	92.56	94.31	92.22
	SGD	89.89	92.56	89.89	91.02
	FBT	**98.10**	**92.49**	**98.10**	**98.00**

(Continued)

Table 3. Comparison of ML and FBT models with different filter feature selection methods *(Continued)*

Feature Selection Method	Model Name	Accuracy	Precision	Recall	F1 Score
Relief F	LR	94.15	88.65	94.15	91.32
	SVM	95.10	94.59	95.10	94.78
	KNN	94.31	92.82	54.31	91.97
	NB	29.23	92.14	29.23	38.81
	SGD	94.94	94.16	94.94	94.38
	FBT	**98.10**	**98.04**	**98.10**	**98.00**

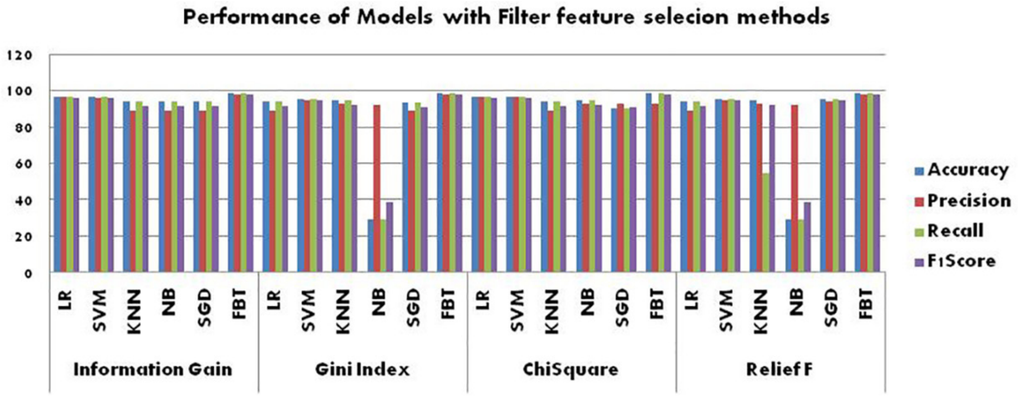

Fig. 2. Performance comparison with filter FS methods

As depicted in Figure 2, information gain and chi-square feature selection methods consistently yield strong performance for LR and SVM models. KNN, NB, and SGD models also benefit from these methods, but to a slightly lesser extent. The Gini, on the other hand, doesn't consistently improve model performance, especially for NB. Relief F is generally effective but can lead to reduced recall for KNN. Notably, FBT consistently outperforms other models across all feature selection methods, demonstrating its robustness and suitability for the given task.

4.2 Experimental results for wrapper and embedded FS methods using UCI data set

Table 4. Comparison of ML and FBT model with different wrapper and embedded feature selection methods

Feature Selection Method	Model Name	Accuracy	Precision	Recall	F1 Score
Forward Selection	LR	94.15	88.65	94.15	91.32
	SVM	94.00	88.64	94.00	91.24
	KNN	95.73	95.32	94.73	95.46
	NB	94.15	88.65	94.15	91.32
	SGD	92.58	90.97	92.58	91.68
	FBT	**97.00**	**96.80**	**97.00**	**96.85**

(Continued)

Table 4. Comparison of ML and FBT model with different wrapper and embedded feature selection methods *(Continued)*

Feature Selection Method	Model Name	Accuracy	Precision	Recall	F1 Score
Backward Selection	LR	94.15	88.65	94.15	91.32
	SVM	94.00	88.64	94.00	91.24
	KNN	95.73	95.32	94.73	95.46
	NB	94.15	88.65	94.15	91.32
	SGD	92.58	90.97	92.58	91.68
	FBT	**97.00**	**96.80**	**97.00**	**96.85**
Recursive Feature Elimination	LR	94.15	88.65	94.15	91.32
	SVM	96.37	96.28	96.37	95.68
	KNN	96.52	96.21	96.52	96.14
	NB	94.15	88.65	94.15	91.32
	SGD	94.15	88.65	94.15	91.32
	FBT	**97.47**	**97.32**	**97.47**	**97.33**
LASSO (with alpha 0.001)	LR	96.84	96.67	96.84	96.43
	SVM	95.42	95.18	95.42	94.15
	KNN	94.15	88.65	94.15	91.32
	NB	95.42	95.98	95.42	95.65
	SGD	96.52	96.21	96.52	96.14
	FBT	**95.58**	**95.02**	**95.58**	**94.68**

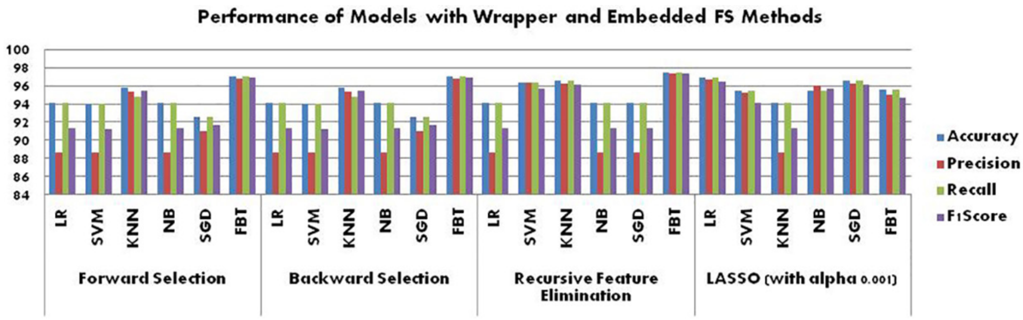

Fig. 3. Performance comparison with wrapper and embedded FS methods

As illustrated in Figure 3, forward selection and backward selection methods tend to maintain consistent performance for LR, SVM, NB, and SGD models, with KNN showing slightly better results. Recursive feature elimination significantly enhances SVM and KNN models' performance, making them stand out. LASSO feature selection also yields improved results for LR and SGD models, with FBT demonstrating competitive performance across all feature selection methods. The choice of feature selection method may depend on the specific requirements of the problem and the selected machine learning model.

4.3 Experimental results for LMH data set

Table 5 presents a comparison of the accuracy of all models with the proposed FBT model using the LMH dataset.

Table 5. Comparison of ML and FBT model with different wrapper and embedded feature selection methods

	All Features	IG	GI	CS	RF	FS	BE	REF	LASSO
SVM	91.30	90.16	93.48	91.30	93.48	95.65	95.65	93.48	95.65
KNN	95.65	95.65	95.65	95.65	95.65	95.65	95.65	95.65	95.65
LR	93.48	93.48	95.65	93.48	93.48	93.48	93.48	93.48	93.48
NB	93.40	93.48	93.48	91.30	93.48	91.30	91.30	93.48	93.48
SGD	93.48	93.48	91.30	91.30	93.48	93.48	95.65	93.48	95.65
FBT	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**97.83**	**97.83**	**100.00**	**97.83**

Figure 4 illustrates the graphical representation of model accuracies, including SVM, KNN, LR, NB, SGD, and FBT, concerning various feature selection methods investigated in the study.

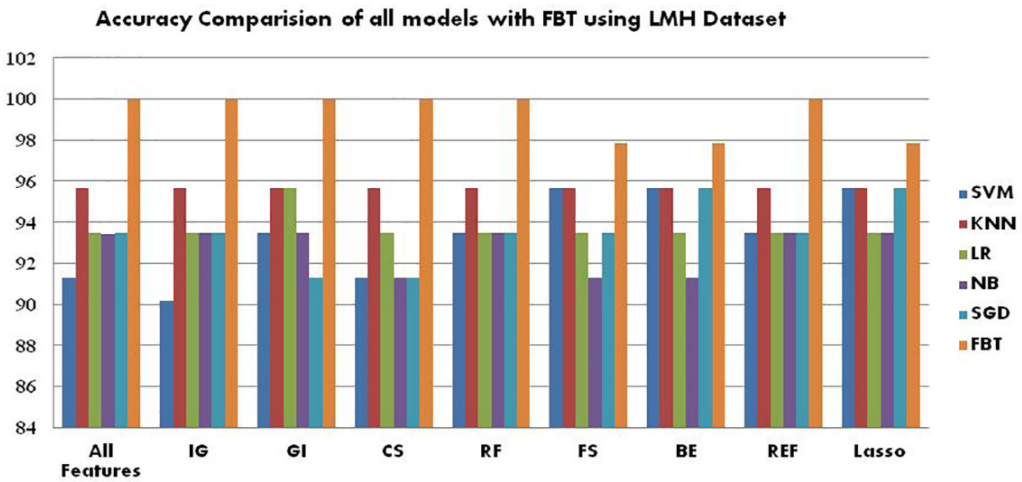

Fig. 4. Accuracy comparison of all models with FBT using LMH dataset

In summary, the FBT model consistently outperforms other models, achieving perfect accuracy in most cases. Feature selection methods play a crucial role, with different models responding differently to each method.

5 CONCLUSION

This paper investigates machine learning algorithms and feature selection strategies for handling classification challenges in depth. Some algorithms show high precision, while others achieve high recall, or F1-score. Specific feature selection strategies have consistently improved model performance, but others have had varied effects depending on the chosen algorithm. Furthermore, a new ensemble model called 'FeatureBoostThyro' (FBT) that combines AdaBoost and grid search optimisation approaches has been introduced. Notably, FBT regularly outperforms other models using different feature selection strategies. Finally, this study has provided insights into the selection of machine learning algorithms and feature selection strategies for solving thyroid condition classification tasks. Again, the development of FBT emphasizes the possibility of improving model performance through novel ensemble tactics. Based on the observations from Figures 3 and 4, it can be inferred that the performance of the FBT model is commendable across both datasets. There is a need to address the unique problems presented by various datasets and problem domains. This study lays a solid foundation for future efforts to improve the effectiveness of machine learning models for classification tasks.

The study has certain limitations that warrant consideration. Firstly, the generalization of the proposed FBT model beyond thyroid disease may be challenging, necessitating further validation for its applicability to different medical conditions. Additionally, the study relies heavily on two datasets—one from Lata Mangeshkar Hospital and another from the UCI dataset—which might limit the model's robustness across diverse populations. The effectiveness of the proposed model is contingent on the quality and representativeness of the datasets, highlighting the importance of rigorous data validation. While the study explores various feature selection methods, there is potential for investigating additional advanced techniques to optimize model performance. Looking ahead, there are promising avenues for future research. Collaborating with medical professionals for clinical validation will enhance the model's credibility in real-world healthcare settings. The inclusion of more diverse datasets covering various demographics, regions, and healthcare systems can further broaden the model's generalizability. Exploring ensemble techniques that combine multiple models could enhance predictive accuracy and overall model robustness. Moreover, efforts to improve the interpretability of the FBT model would facilitate its adoption by healthcare practitioners in clinical decision-making. Longitudinal studies incorporating temporal data can assess the model's effectiveness in predicting disease progression over time. Finally, investigating the feasibility of integrating the FBT model into existing healthcare systems is essential for its seamless adoption in routine clinical practice. These future directions aim to address limitations and contribute to the advancement and practical application of machine learning techniques in early-stage disease prediction and healthcare decision-making.

6 REFERENCES

[1] M. King, X. Ma, B. Xi, Y. Zhang, L. Zhu, S. Xin, G. Tian, and J. Yang, "A machine learning-based diagnosis of thyroid cancer using thyroid nodules ultrasound images," *Current Bioinformatics,* vol. 15, no. 4, pp. 349–358, 2020. https://doi.org/10.2174/1574893614666191017091959

[2] K. B. Raju, P. K. Lakineni, K. S. Indrani, G. M. S. Latha, and K. Saikumar, "Optimized building of machine learning technique for thyroid monitoring and analysis," in *2nd International Conference on Smart Electronics and Communication (ICOSEC),* 2021, pp. 1–6. https://doi.org/10.1109/ICOSEC51865.2021.9591814

[3] S. Nandhinidevi, S. Poorani, and P. G. Brindha, "Machine learning models for relevant feature identification and classification of thyroid data," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 9, no. 5, pp. 1961–1963, 2020. https://doi.org/10.35940/ijitee.E2948.039520

[4] A. K. Chaudhuri, D. K. Banerjee, A. Das, and A. Ray, "A multi-stage approach combining feature selection with machine learning techniques for higher prediction reliability and accuracy in heart disease diagnosis," *International Journal of Engineering Research & Technology (IJERT)*, vol. 10, no. 7, pp. 73–85, 2020. https://dx.doi.org/10.17577/IJERTV10IS070057

[5] K. V. V. Reddy, I. Elamvazuthi, A. A. Aziz, S. Paramasivam, H. N. Chua, and S. Pranavanand, "Heart disease risk prediction using machine learning classifiers with attribute evaluators," *Applied Sciences*, vol. 11, no. 18, p. 8352, 2021. https://doi.org/10.3390/app11188352

[6] K. Juneja, "Expanded and filtered features based ELM model for thyroid disease classification," *Wireless Personal Communications*, vol. 126, pp. 1–38, 2022. https://doi.org/10.1007/s11277-022-09823-7

[7] R. Varatharajan, G. Manogaran, and M. K. Priyan, "A big data classification approach using LDA with an enhanced SVM method for ECG signals in cloud computing," *Multimedia Tools and Applications*, vol. 77, pp. 10195–10215, 2017. https://doi.org/10.1007/s11042-017-5318-1

[8] G. Chaubey, D. Bisen, S. K. Arjaria, and V. Yadav, "Thyroid disease prediction using machine learning approaches," *National Academy Science Letters*, vol. 44, no. 3, pp. 233–238, 2021. https://doi.org/10.1007/s40009-020-00979-z

[9] P. Duggal and S. Shukla, "Prediction of thyroid disorders using advanced machine learning techniques," in *10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2020, pp. 670–675. https://doi.org/10.1109/Confluence47617.2020.9058102

[10] K. Shankar, S. K. Lakshmanaprabu, D. Gupta, A. Maseleno, V. H. C. de Albuquerque, "Optimal feature-based multi-kernel SVM approach for thyroid disease classification," *The Journal of Supercomputing*, vol. 76, pp. 1128–1143, 2018. https://doi.org/10.1007/s11227-018-2469-4

[11] R. Quinlan, "Thyroid disease," *UCI Machine Learning Repository*, 1987. https://doi.org/10.24432/C5D010

[12] A. R. Rao and B. S. Renuka, "A machine learning approach to predict thyroid disease at early stages of diagnosis," in *IEEE International Conference for Innovation in Technology (INOCON)*, 2020, pp. 1–4. https://doi.org/10.1109/INOCON50539.2020.9298252

7 AUTHORS

Deepali Bhende has earned a Master's degree in Computer Science from G. H. Raisoni Institute of Information Technology, Nagpur. Presently, she is actively pursuing her Ph.D. at G. H. Raisoni University, Saikheda (MP). Over her impressive 18-year career in education, she has actively participated in numerous National and International conferences. Beyond her extensive involvement in academia, she has made notable contributions through several published papers in various international journals, with two papers indexed in SCOPUS. Furthermore, she has successfully obtained a patent, demonstrating her innovative contributions to the field (E-mail: deepali.bhende.phdcs@ghru.edu.in).

Dr. Gopal Sakarkar is an experienced academic, holds a Master's degree (2006) and a Ph.D. (2017) from S.G. B. Amravati University, Amravati, accumulating over 15 years of teaching and research expertise. Currently he is working as an Associate Professor at Dr. Vishwanath Karad MIT World Peace University, Pune, Dr. Sakarkar has demonstrated a commitment to education by shaping curriculum and schemes as a BOS member of the Department of AI and Data Science at GHRCE, Nagpur. He is also serving as an External-Academic Board of Study Member at Government

Polytechnic College, Nagpur, and HVPM College, Amravati and has contributed to syllabus design. As a research guide, he mentors Ph.D. scholars, and his prolific research output comprises 55+ papers in international journals and conferences, including SCI, ESCI, WoS, Springer, Elsevier, IEEE, AIP, and SCOPUS indexed conferences. With 200+ Google Scholar citations, he holds a 6 h-index, a 5 i10 index, and 79 citations in Scopus with a 3 h-index.

Punam Khandar has completed her M. Tech(CSE) degree from RTMNU, Nagpur, in the year 2015 and currently serves as an Assistant Professor at Shri Ramdeobaba College of Engineering and Management, Nagpur. She is presently pursuing a Ph.D. in Computer Science and Engineering at KIIT, Deemed to be University, Bhubaneswar. Her research interests encompass Machine Learning, Deep Learning, and Computer Vision. She has recently authored papers on various topics, including a survey on Decision Trees, detection of leaf diseases, identification of heart diseases, and the use of X-rays for identifying COVID-19.

Satyajit Uparkar is a certified data scientist, has been an Assistant Professor at Shri Ramdeobaba College of Engineering and Management, Nagpur, for the past 11 years. Holding three postgraduate degrees, he specializes in data analytics with a focus on Data Mining, Scalable Data Science, and Operation Research Modeling. He has received two best paper awards at international conferences and has contributed 30 research papers to various national and international journals. Alongside his academic role, he serves as a data science consultant for local companies.

Arvind Bhave is an Assistant Professor in the Department of Electronics and Computer Science at R.T.M. Nagpur University, Nagpur. He earned his B.E. in Computer Science from Amravati University and M. Tech in Computer Science and Engineering from RTM Nagpur University in 2014. Currently, he is pursuing a Ph.D. in Computer Science and Engineering from SGB Amravati University, Amravati. His research interests include Image Processing, Computer Vision, Machine Learning, and Video Processing.