CHAPTER 4

FEATURE SELECTION BY IMPORTANCE SCORES OF FEATURES FOR HEART DISEASE PREDICTION

This chapter discusses the adoption of Feature Importance (FI) scores by gradient boosting algorithms, CatBoost (CatB) and XGBoost, for attribute selection. Applying the forward selection method yields subsets of features using the FI ranking as threshold. The classifier models XGBoost, Hard Maj.Vote and CatB are used for classification. The optimal set of characteristics which yields the highest accuracy and the range in which the optimal set of attributes can be found is identified which subsequently reduces the number of searches to be done.

4.1 INTRODUCTION

4.1.1 Feature Importance (FI)

Feature Importance (FI) or Importance Score of features enables us to comprehend the importance of each input feature while making predictions of the target variable. In tree models, the structure of the tree and how it divides the data into different aspects usually determine which features are important to the model. The two most popular ways to determine FI are as follows:

i. Gini importance:

The relevance of a feature for splitting purposes is quantified using the Gini index. The model's trees are traversed in order to determine the aggregate impurity reduction at each node splitting on that feature. More weight is given to features with a higher Gini relevance in the model's predictions.

ii. Permutation importance:

Permutation importance assesses the impact of shuffling the values of a feature on the model's performance. It measures the decrease in model performance (such as accuracy or MSE) when a feature's values are randomly permuted. If a feature is important, permuting its values will result in a significant drop in the model's performance. Both methods provide a relative measure of FI. Higher values indicate greater importance, while lower values indicate lesser importance. It's important to note that FI in tree models is not always directly related to the magnitude or correlation of the feature with the target variable, as tree models consider interactions among features.

FI in tree models can be put to use in a number of contexts, including but not limited to prediction, feature selection, data exploration, and anomaly/noise detection. However, it's important to interpret FI in the context of the specific model and dataset being used.

4.1.2 Feature Selection Using Importance Score of Features

In ML, the dimensionality of the input features can be decreased through feature selection utilizing feature priority ranking. The goal is to prioritize features and use just the most important ones in the predictive model. This process involves the following steps:

1. Train a tree-based model: First, train a tree-based model on the training data. Some examples include DT, RF, XGBoost, and CatB. These models inherently calculate FI during the training process.
2. Retrieve FI scores: Extract the FI scores from the trained model. The importance scores can be obtained using methods specific to the tree-based model used, such as Gini importance or permutation importance.
3. Rank the features: Arrange the features from most important to least. This ranking provides an indication of the relative significance of each feature in predicting the target variable.
4. Select the top features: Set a minimum or target number of features to retain. Select the top-ranked features that meet the criteria. Advanced feature selection techniques like Recursive Feature Elimination (RFE) or forward/backward stepwise selection can be used.
5. Build the model with selected features: Finally, build the predictive model using only the selected features.

The low-ranking features from earlier, similar works were omitted based on experimental results and the dataset used at the time.

By means of FI by gain in XGBoost, Din. H et al. (2019) picked the best 24 features from the NHANES dataset and discarded the remainder due to poor performance.

Using Infogain, Correlation, and ReliefF, Zahangir et al. (2019) selected features for the Statlog dataset; ReliefF was scored lowest, hence it was left out of the final subset. The top two features from the feature ranking by RF using Classifier Attribute were omitted from the final subset of the SA heart dataset.

Using IG, reliefF, gain ratio, oneR, and symmetric uncertainty, Mohammad Ashraf Ottom and Walaa Alshorman (2019) ranked features from the Cleveland heart dataset. Elimination of low-ranking features.

These works suggest that the experimental findings were used to randomly delete the low performing features. In this study, XGBoost and CatB, two popular gradient boosting algorithms, were used to get the FI values of the attributes in a dataset. Each FI value was used as a threshold to choose features in a forward selection process, where features were prioritized based on their rank. The subsets were used to train and compare various classifiers, XGBoost, CatB and MVE (which consists of LR, RF, GNB, XGBoost, and CatB). Instead of exhaustive search done across the subsets of features formed based onthe FI values, a range in which the best subset of features can be found is identified in this work. This would reduce the search space for feature subsets when compared to exhaustive search.

4.1.3 Machine Learning Classifiers

4.1.3.1 EXtreme Gradient Boosting (XGBoost)

XGBoost, which is gradient-boosted DTs (Chen & Guestrin, 2016), requires less training time. In XGBoost, objective function, is the sum of training loss and regularization, where training loss is the differentiable convex loss function 'L' which computes the difference between the target y_i and the prediction $\hat{y}_i$. The regularization term is added in order to smooth the final learnt weights and thereby prevent over-fitting. The objective function is determined as follows:

$$Obj^{i} = \sum_{i=1}^{n} L\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{i=1}^{t} \Omega(f_i) \tag{4.1}$$

In Eq. (4.1), $\hat{y}_i^{(t)}$ is the prediction at t^{th} round, a DT is represented by the structure, f_i and to which the regularization term $\Omega(f_i)$ is associated, is defined as:

$$\Omega(f_i) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{t} \omega_j^2 \tag{4.2}$$

In Eq. (4.2), ‘T’ represents the number of leaf nodes in the current tree ‘f_i’. The term ‘γT’ directly penalizes the complexity of the tree based on the number of leaves. ‘ω_j’ represents the prediction (score) assigned to the j-th leaf node in the tree. ‘λ’ is the L2 regularization term weight applied to the leaf weights ω_j.

By minimizing the training loss, models would perform better. Regularization is a tool for gauging a model's complexity. The model is simplified using regularization optimization.

In XGBoost, the various feature importance categories are weight, gain, cover, total_gain, and total_cover, where the default type is weight. Weight is the total number of trees in which that feature was used to create a partition. Gain is determined as the average gain value of all splits where the feature is used. The cover is equal to the weighted average of all splits in which the feature is enabled. When a feature is used in many splits, the aggregate gain for those splits is added together to form total_gain. The feature's total coverage, or total_cover, includes all of the splits in which it was used.

4.1.3.2 CatBoost (CatB)

CatB, a gradient boosting implementation, was created by Yandex scientists and engineers. CatB 's base predictors are binary DT’s. In order to use this algorithm, all that has to be adjusted is the learning rate and the number of iterations. CatB is an application of ordered boosting based on a random permutation technique. It efficiently represents categorical data as a vector, making it suitable for datasets with many categorical features (Prokhorenkova et al., 2018). CatB outperforms XGBoost if the hyper parameters are not tuned and only the default parameters are used. CatB's CPU and GPU implementations are both orders of magnitude quicker than competing solutions.

CatB has feature importance types: ‘Prediction Values Change (PVC)’ and ‘Loss Function Change (LFC)’. PVC is the average changes that occur in the prediction when a change occurs in the value of a feature. To determine a feature's LFC, one must compare the model’s loss with and without the inclusion of the feature.

4.1.3.3 Majority Vote Ensemble

ML model "MVE" aggregates predictions from multiple models. The ensemble of models' predictions for a given class label are added up in a process called "hard voting”, and the most popular categorization is chosen.

The FI aspect of XGBoost and CatB is exploited to perform feature selection in this study. Both of these classifiers, together with the Hard MVE comprising LR, RF, Gaussian Nave Bayes, XGBoost, and CatB, are modeled on datasets related to cardiovascular illness, and their performance metrics are compared and contrasted.

4.2 METHODOLOGY

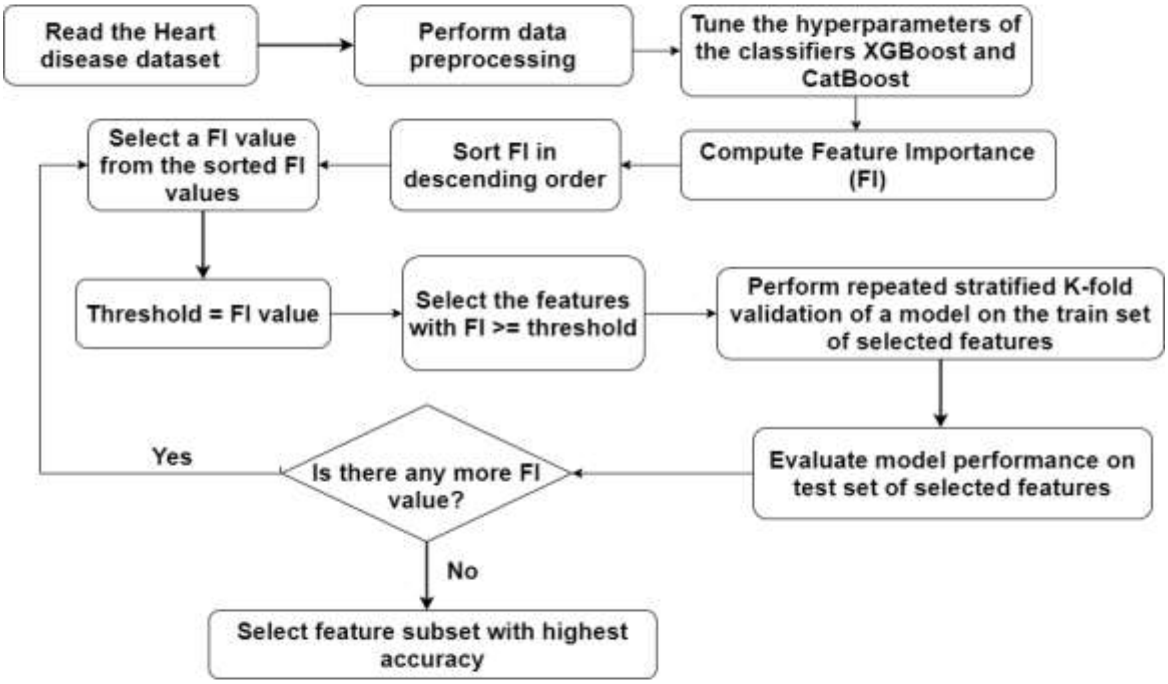

Figure 4.1 Methodology for FI based Selection of Features and Classification on Heart Disease Dataset

Feature Importance type ‘Prediction Values Change’ of CatB and ‘gain’ of XGBoost were investigated on South African (SA) heart, Statlog (Stat) heart, Z-Alizadeh Sani (Z-Ali) heart, and Cleveland (Clev) heart dataset (described in section 3.4). Each FI value was used as a threshold to choose features in a forward selection process, where features were prioritized based on their rank. Classifiers like XGBoost and CatB were trained on subsets of the dataset, as was the Hard MVE Classifier (comprising GNB, RF, XGBoost, LR, and CatB). The best set of features was determined by their ability to produce a high degree of accuracy.

The range of FI scores where the optimal feature subset yielding the highest accuracy would be found was identified and validated across the four heart datasets.

4.3. Analysis of FI Score types of Gradient Boosting Algorithms on Heart Datasets

In all the datasets, initially missing values are filled, then numeric values are normalized using the Min-Max scaling function mentioned in Eq.4.3.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{4.3}$$

The FI Score by PVC type and gain type, of CatB and XGBoost respectively, were done on all four heart datasets mentioned in section 4.2. Figure 4.2 and Figure 4.3 depict importance scores of features by gain type and importance scores of features by PVC type, on SA heart dataset. We can observe that the top 5 characteristics selected by both methods are the same.

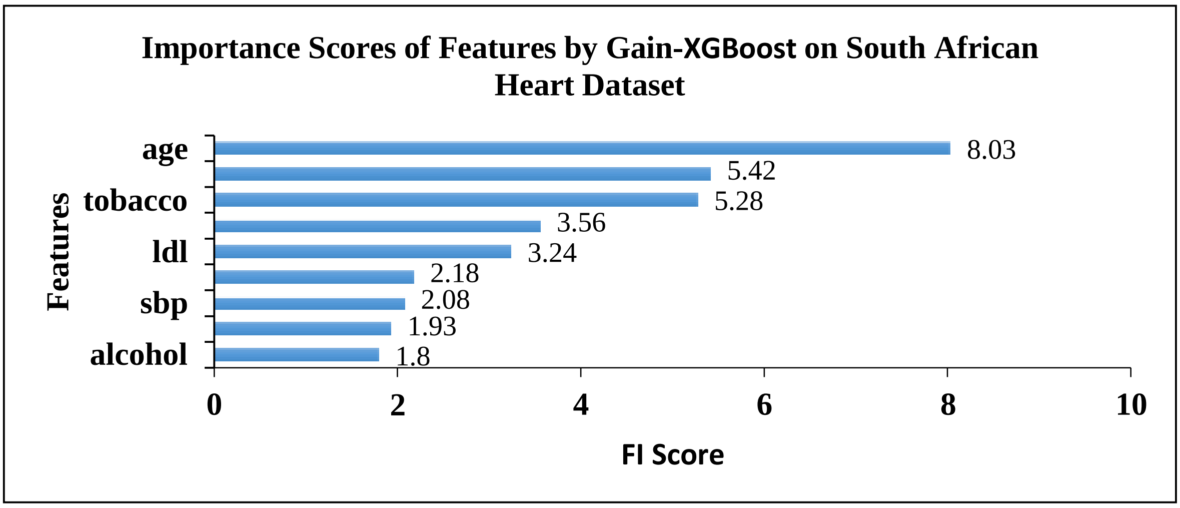

Figure 4.2 Importance Scores of Features by Gain- XGBoost on SA Heart Dataset

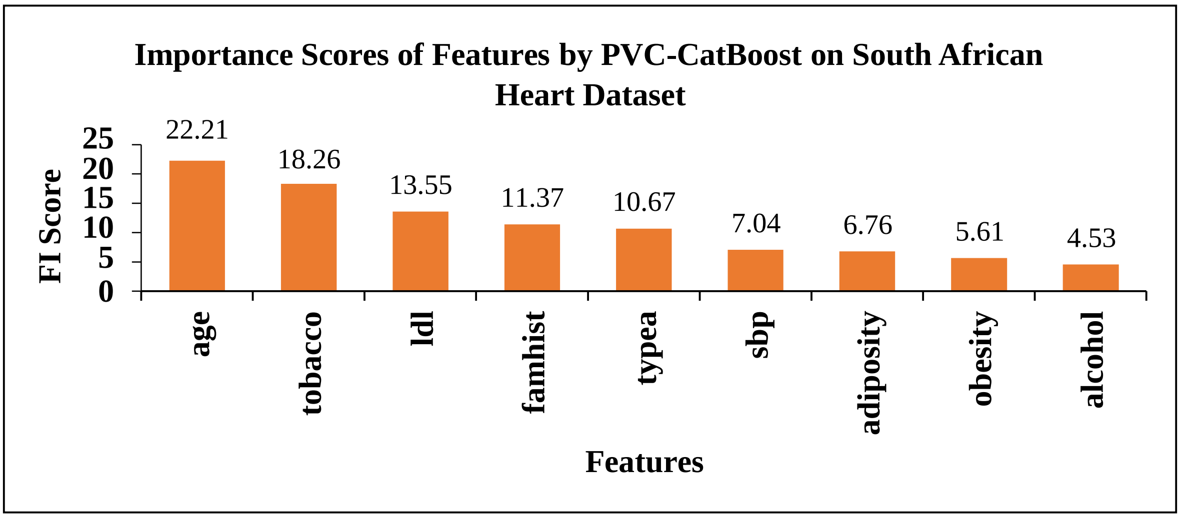

Figure 4.3 Importance Scores of Features by PVC-CatBoost on SA Heart Dataset

Figure 4.4 and Figure 4.5 display gain type importance score and PVC type importance score on Cleveland heart dataset, which show the top three features are similar.

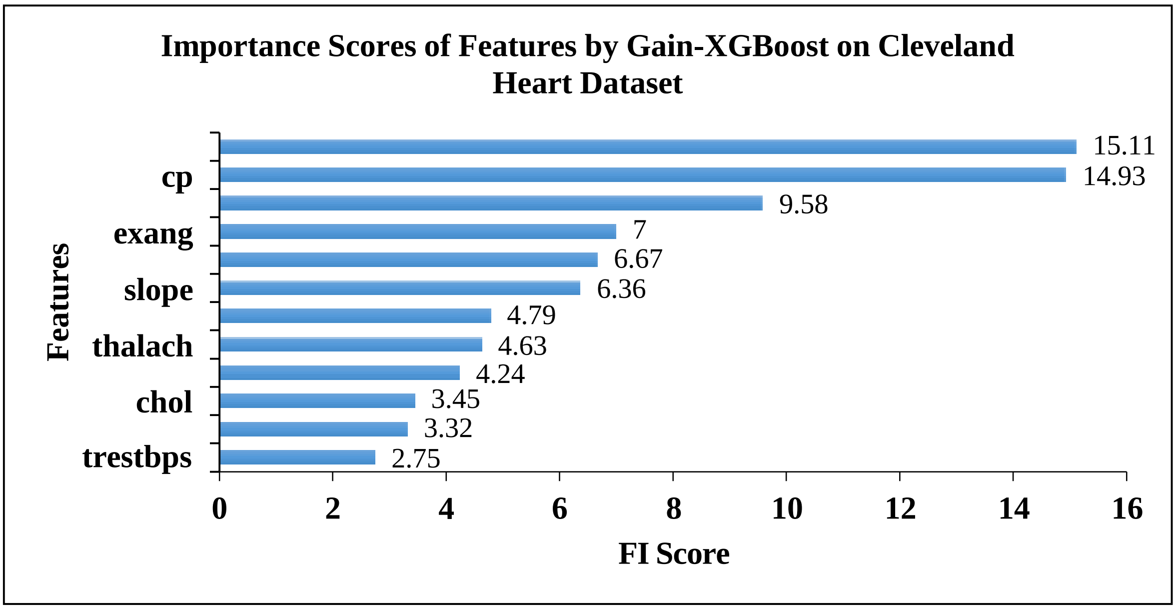

Figure 4.4 Importance Scores of Features by Gain-XGBoost on Cleveland Heart Dataset

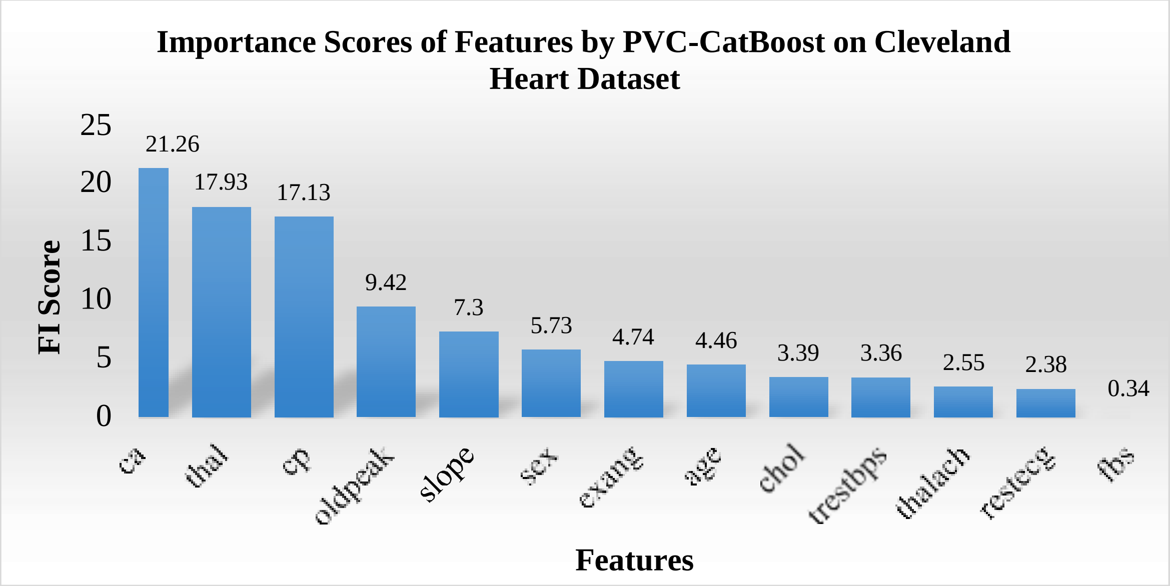

Figure 4.5 Importance Scores of Features by PVC-CatBoost on Cleveland Heart Dataset

Figure 4.6 and Figure 4.7 display the FI by gain of XGBoost and FI by PVC of CatBoost on Statlog heart disease dataset, where in both the rankings, the top 3 features are same.

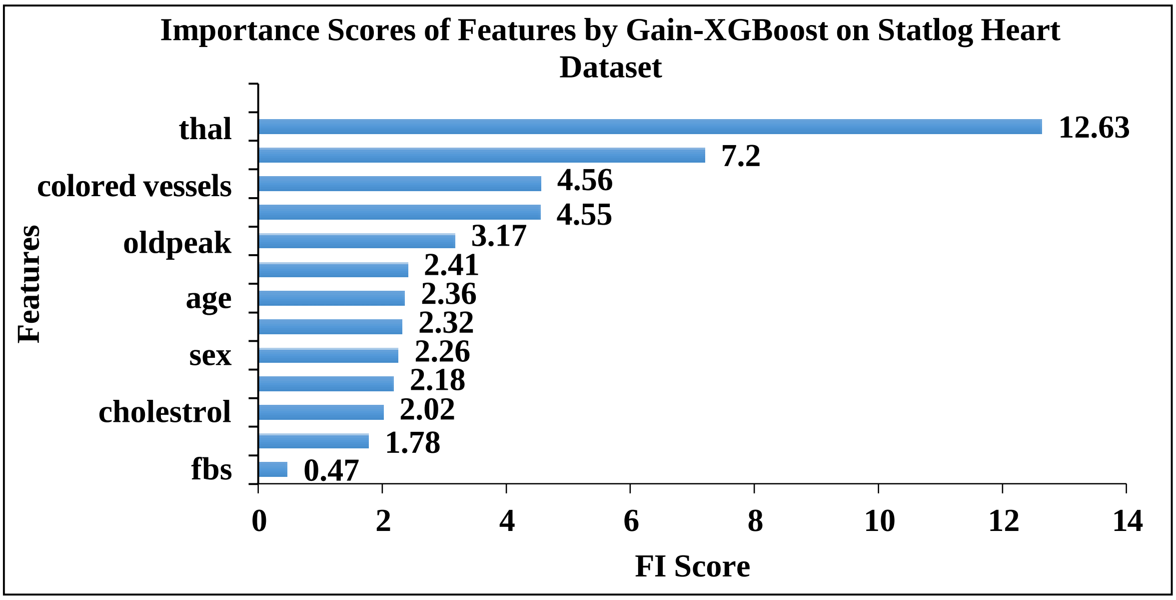

Figure 4.6 Importance Scores of Features by Gain-XGBoost on Statlog Heart Dataset

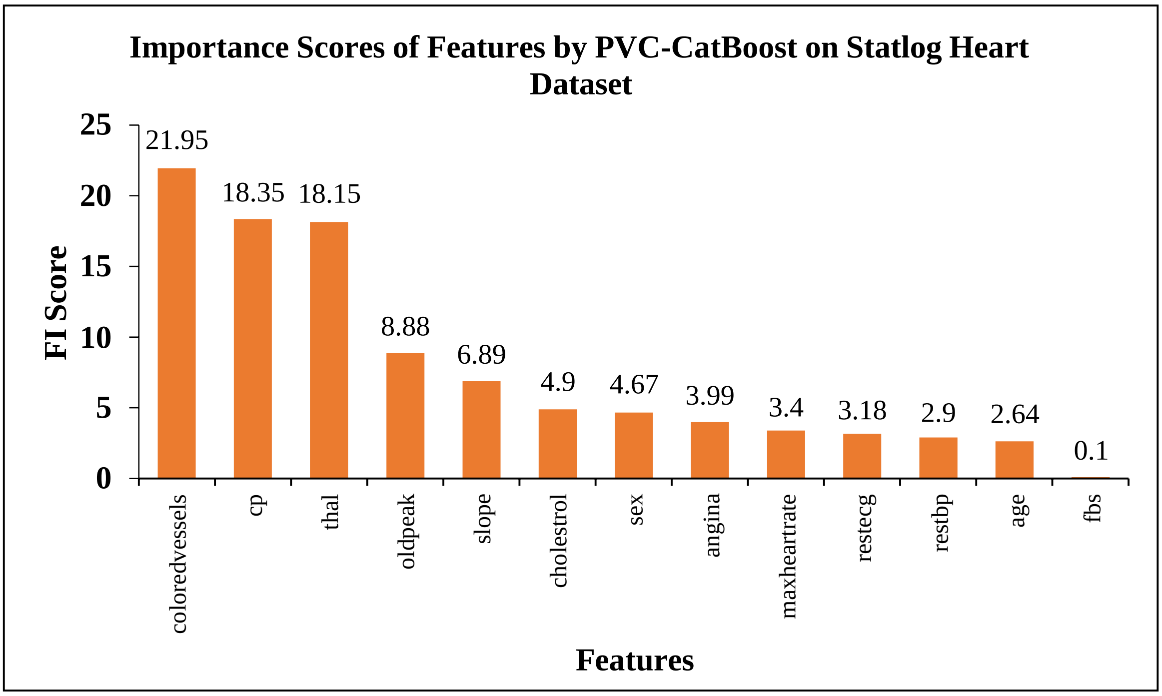

Figure 4.7 Importance Scores of Features by PVC-CatB on Statlog Heart Dataset

Applying FI Score on Z-Ali heart dataset, a high dimensional dataset consisting of 55 predictor features (described in section 3.4), the top 30 features based on the FI Score by gain type is displayed in the Table 4.1. These features have FI > 0.5.

Table 4.1. Importance Scores of Features by Gain-XGBoost on Z-Alizadeh Sani Heart Dataset

S.no	Feature	FI Score	S.no	Feature	FI Score
1	Chest Pain	**44.642**	16	WBC	2.769
2	HTN	**12.462**	17	HDL	**2.553**
3	Region RWMA	**9.877**	18	Non anginal	**2.464**
4	Tin version	**7.287**	19	FBS	**2.397**
5	Age	**6.967**	20	PR	**2.35**
6	DM	**6**	21	PLT	**2.328**
7	Atypical	**5.484**	22	Weight	**2.192**
8	EF-TTE	**5.185**	23	LDL	**1.925**
9	Cur-Smoker	**4.74**	24	ESR	**1.916**
10	CR	**3.843**	25	Na	**1.757**
11	FH	**3.393**	26	BP	**1.752**
12	K	**3.256**	27	Dyspnea	**1.516**
13	TG	**3.251**	28	HB	**1.416**
14	BMI	**2.935**	29	Neut	**1.356**
15	ST Depression	**2.78**	30	Lymph	**0.786**

From Figure 4.8, which shows the FI Scores of the features by Gain type and Figure 4.9 which shows the FI Scores of the features by PVC type, it is very clear that the feature chest pain highly contributes to the prediction of heart disease. This is followed closely by hypertension and then by Region Wall Motion Abnormality Region (RWMA) and age. This indicates that a person cannot ignore chest pain as it might be a threat to his/her life. In Figure 4.9, there are 17 features with values < 0.1 which may not significantly contribute to the prediction of heart disease.

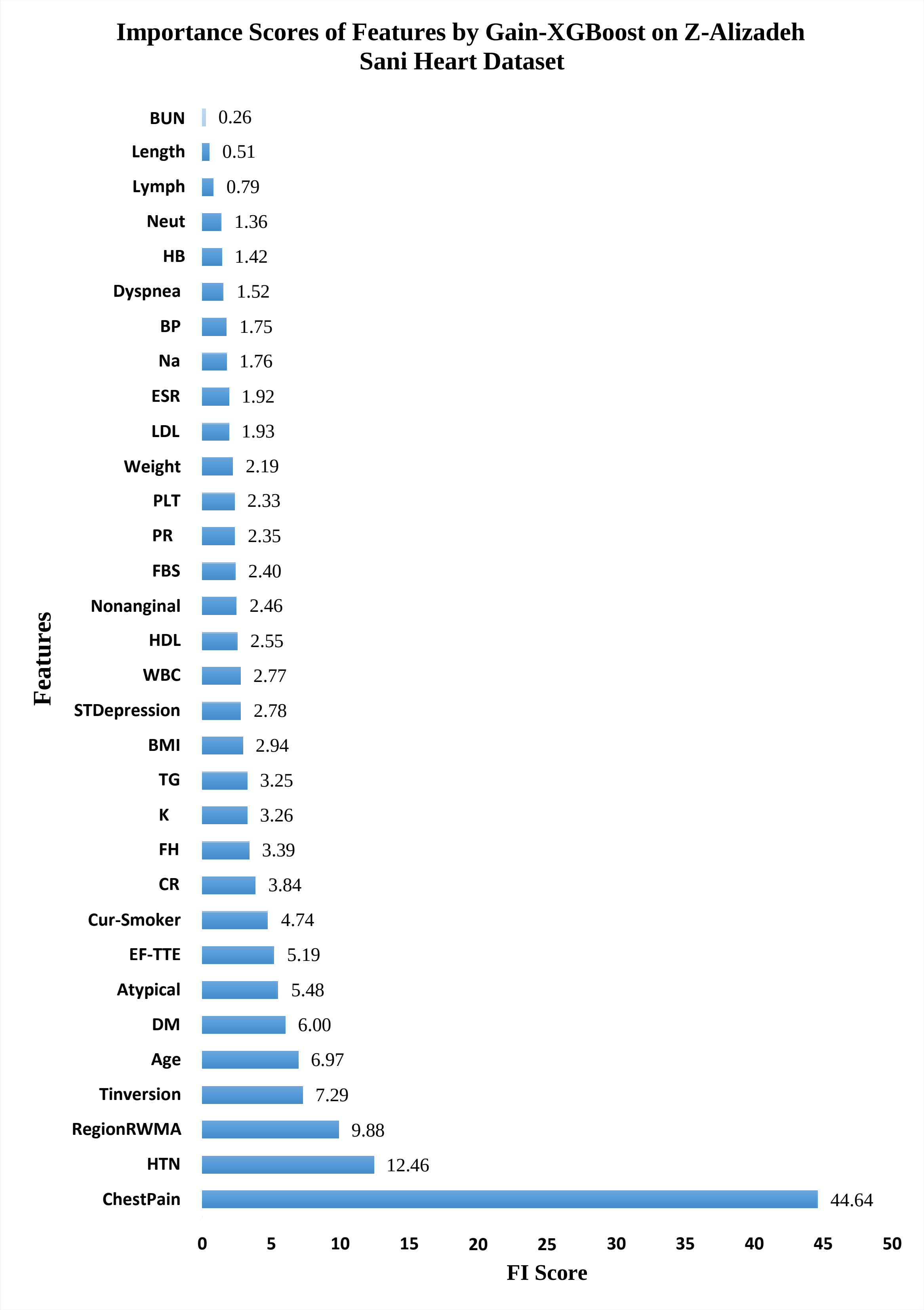

Figure 4.8 Importance Scores of Features by Gain-XGBoost on Z-Ali Heart Dataset

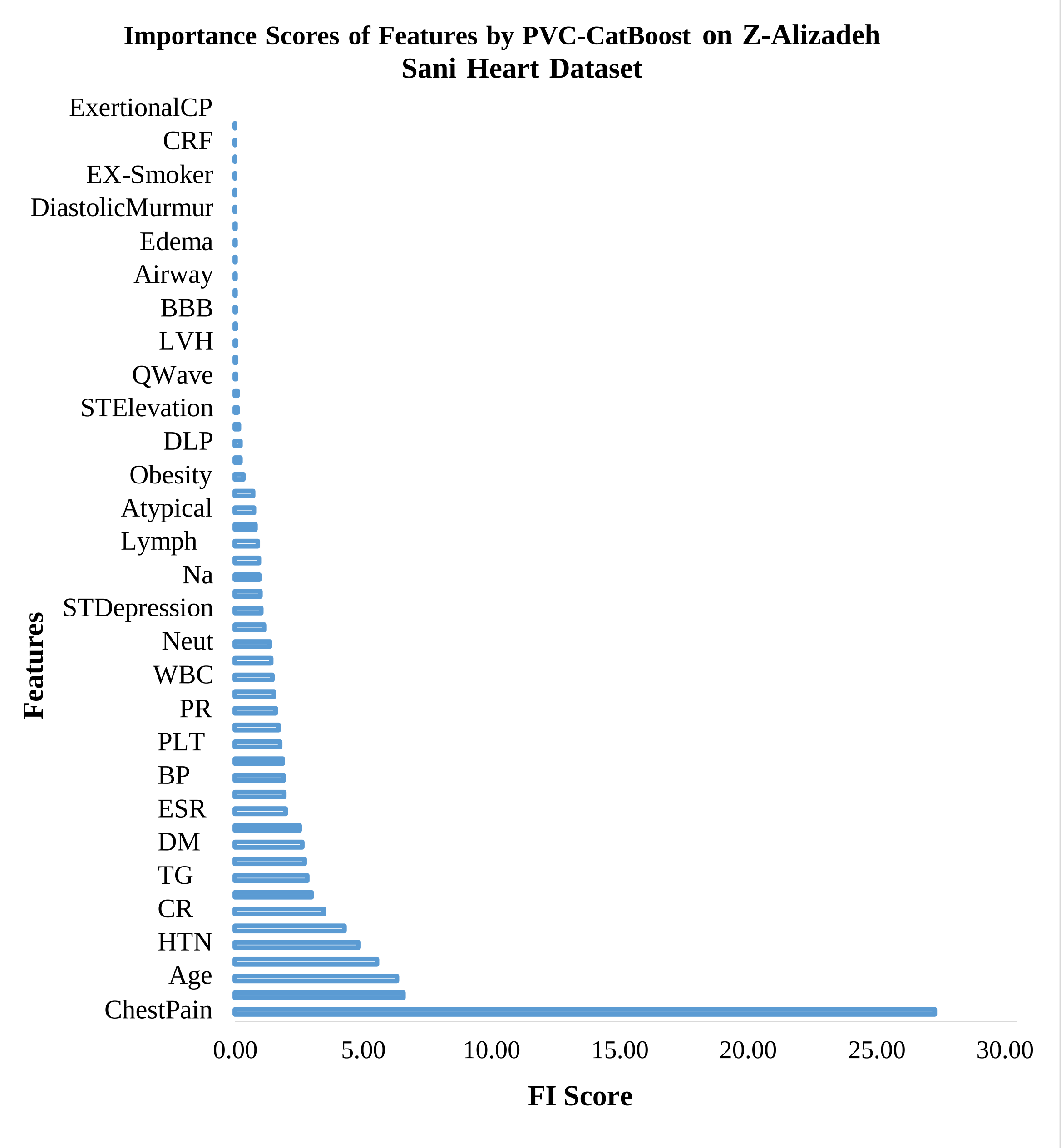

Figure 4.9 Importance Scores of Features by PVC-CatB on Z-Ali Heart Dataset

The Importance Scores of Features by PVC on Z-Ali heart dataset is shown in Table 4.2. It also shows that the feature 'chest pain' is the leading predictor of heart disease. It is known that features such as edema, thyroid, chest pain due to exertion are notrisk factors and accordingly the CatBoost- PVC type ranks them low. So, it is shown that more features which have very less significance towards prediction of the disease are present and it would be ideal to do feature selection so that the model would perform in a much better way.

Table 4.2. Importance Scores of Features by CatBoost-PVC on Z-Alizadeh Sani Heart Dataset

S.No	Features	FI Score	S.No	Features	FI Score
1	ChestPain	27.294072	29	Lymph	0.891004
2	RegionRWMA	6.569491	30	VHD	0.803514
3	Age	6.320182	31	Atypical	0.738841
4	EF-TTE	5.531393	32	FunctionClass	0.707019
5	HTN	4.824725	33	Obesity	0.323713
6	Tinversion	4.270848	34	Sex	0.210267
7	CR	3.466479	35	DLP	0.209549
8	BMI	2.98846	36	Dyspnea	0.150642
9	TG	2.815079	37	STElevation	0.101795
10	FBS	2.715833	38	Nonanginal	0.100949
11	DM	2.626709	39	QWave	0.044873
12	K	2.518097	40	PoorRProgression	0.037058
13	ESR	1.975211	41	LVH	0.036247
14	HDL	1.923607	42	CVA	0.033367
15	BP	1.899305	43	BBB	0.026806
16	HB	1.859965	44	Lung-rales	0.023186
17	PLT	1.759165	45	Airway	0.018249
18	Weight	1.710643	46	SystolicMurmur	0.016121
19	PR	1.593187	47	Edema	0.015924
20	LDL	1.528747	48	Thyroid	0.012008
21	WBC	1.461806	49	DiastolicMurmur	0.006604
22	FH	1.409719	50	CHF	0.003591
23	Neut	1.370001	51	EX-Smoker	0.003103
24	Cur-Smoker	1.154062	52	LowTHAng	0.002091
25	STDepression	1.023721	53	CRF	0.001253
26	BUN	0.990238	54	WeakPeripheralPulse	0.000739
27	Na	0.944003	55	ExertionalCP	0
28	Length	0.936738			

4.4 Evaluation of Performance Accuracy of the Classifiers

The datasets were split in the ratio 80:20 for training and testing. Subsets of features were formed by taking each FI score as threshold, beginning with the highest score. Classifiers XGBoost, CatB and MVE were modeled on these subsets using 5-fold Stratified cross-validation with three repeats. The maximum accuracy yielding subset isthe optimal one for each dataset. Since the most highly rated features are prioritized, the resulting feature subsets are the best possible choices.

Table 4.3 – Table 4.6 display the performance accuracy of the classifiers MVE, CatB and XGBoost on each subset of features generated by the PVC and gain FI types of CatB and XBGoost on the different heart datasets. Table 4.3 shows that the highest accuracy exhibited by MVE-gain is 88.52%, MVE-PVC is 90.16%, CatBoost is 91.8% and XGBoost exhibited 88.52%. So, CatBoost classifier modelled on subsets formed by FI type PVC of CatBoost gives the highest accuracy on Clev Heart Dataset. Table 4.4 shows MVE classifier exhibiting high performance accuracy of 87.04% compared to other classifiers on Stat heart dataset.

Table 4.3 Performance Accuracy of Classifiers on Feature Subsets of Cleveland Heart Dataset

Feature Subsets	MVE-Gain	MVE-PVC	CatBoost-PVC	XGB- Gain
Top 1	78.69%	75.41%	75.10%	78.69%
Top 2	78.69%	83.61%	81.97%	78.69%
Top 3	86.89%	85.25%	86.89%	86.69%
Top 4	**88.52%**	**90.16%**	90.16%	**88.52%**
Top 5	85.25%	88.52%	**91.80%**	83.61%
Top 6	86.89%	86.89%	90.16%	85.25%
Top 7	86.89%	86.89%	88.52%	85.25%
Top 8	86.89%	88.52%	88.52%	85.25%
Top 9	86.89%	88.52%	88.52%	83.61%
Top 10	88.52%	88.52%	88.52%	86.89%
Top 11	85.25%	85.25%	85.25%	85.25%
Top 12	81.97%	85.25%	86.89%	85.25%
All 13	81.97%	86.89%	86.89%	85.25%

Table 4.4 Performance Accuracy of Classifiers on Feature Subsets of Statlog Heart Dataset

Feature Subsets	MVE-Gain	MVE-PVC	CatBoost-PVC	XGB- Gain
Top 1	77.78%	72.22%	72.22%	74.075
Top 2	72.22%	66.67%	68.52%	70.37%
Top 3	77.78%	**87.04%**	**85.19%**	72.22%
Top 4	85.19%	81.48%	81.48%	**85.19%**
Top 5	79.63%	83.33%	79.63%	83.33%
Top 6	83.33%	83.33%	79.63%	83.33%
Top 7	85.19%	81.48%	79.63%	83.33%
Top 8	83.33%	83.33%	81.48%	81.48%
Top 9	83.33%	85.19%	83.33%	83.33%
Top 10	**87.04%**	83.33%	83.33%	81.48%
Top 11	85.19%	85.19%	81.48%	81.48%
Top 12	85.19%	85.19%	81.48%	79.63%
Top 13	85.19%	83.33%	83.33%	77.78%

Table 4.5 displays the performance accuracies of the three classifiers modelled on the feature subsets of SA heart dataset. Here, the highest accuracy of 78.49% is given by CatBoost classifier compared to other classifiers.

Table 4.5 Performance Accuracy of classifiers on Feature Subsets of SA Heart Dataset

Feature Subsets	MVE-Gain	MVE-PVC	CatBoost-PVC	XGB- Gain
Top 1	66.67%	68.82%	77.74%	67.74%
Top 2	65.59%	69.89%	72.04%	64.52%
Top 3	70.97%	74.19%	72.04%	72.04%
Top 4	**73.12%**	74.19%	74.19%	69.89%
Top 5	69.89%	**75.27%**	77.42%	**72.04%**
Top 6	68.82%	75.27%	**78.49%**	69.89%
Top 7	69.89%	75.27%	76.34%	72.04%
Top 8	68.82%	74.19%	76.34%	69.89%
All 9	69.89%	75.27%	75.27%	68.82%

Table 4.6 displays the performance accuracies of the classifiers modelled on Z-Ali heart dataset. The feature subsets with FI values close to 0 (less than 0.5) are not considered. According to FI type by XGBoost-Gain, the FI median value is 2.77 and by CatBoost –PVC the FI median value is 1.81. Table 4.6 shows that the highest accuracy is 96.55% given by MVE classifier.

Table 4.6 Performance Accuracy of Classifiers on Feature Subsets of Z-Alizadeh Sani Heart Dataset

Feature Subsets with FI value >0.5	XGB-Gain	MVE-Gain	MVE-PVC	CB-PVC
Top 1	85.06%	83.91%	83.91%	83.00%
Top 2	85.06%	83.91%	90.80%	87.00%
Top 3	91.95%	90.8%	88.51%	89.00%
Top 4	94.25%	93.1%	89.66%	89.00%
Top 5	94.25%	93.1%	91.95%	90.00%
Top 6	94.25%	93.1%	93.10%	90.00%
Top 7	94.25%	93.1%	93.10%	91.00%
Top 8	91.95%	91.95%	94.25%	92.00%
Top 9	91.95%	93.1%	93.10%	92.00%
Top 10	93.1%	94.25%	93.10%	92.00%
Top 11	94.25%	95.55%	93.10%	92.00%
Top 12	91.95%	95.4%	93.10%	91.00%
Top 13	91.96%	94.25%	93.10%	92.00%
Top 14	**95.4%**	95.4%	**95.40%**	92.00%
Top 15	94.25%	**96.15%**	94.25%	**93.00%**
Top 16 (Gain Median FI=2.77)	94.25%	**96.15%**	94.25%	92.00%
Top 17 (PVC Median FI=1.81)	**95.4%**	95.4%	**95.40%**	92.00%
Top 18	94.25%	95.4%	**95.40%**	92.00%
Top 19	95.4%	94.25%	93.10%	92.00%
Top 20	95.4%	95.4%	94.25%	92.00%
Top 21	94.25%	96.15%	93.10%	91.00%
Top 22	94.25%	96.15%	93.10%	92.00%
Top 23	94.25%	95.4%	94.25%	92.00%
Top 24	95.4%	96.15%	94.25%	92.00%
Top 25	95.4%	95.4%	95.40%	92.00%
Top 26	93.1%	95.4%	94.25%	92.00%
Top 27	94.25%	93.1%	93.10%	92.00%
Top 28	94.25%	94.25%	94.25%	92.00%
Top 29	94.25%	96.15%	91.95%	92.00%
Top 30	95.4%	93.1%	93.10%	92.00%

From Tables 4.3- 4.6, it can be noticed that that the feature subsets giving higher accuracies are present around the median FI value. Moreover, the highest accuracy is given by feature subsets which have FI values within 4 FI values above the median.

4.5 Proposed Search Space of Feature Subsets

From Table 4.3 – Table 4.6, it can be noted that the maximum accuracy in each dataset, is given by the subset with has FI threshold value close to the median FI value. The range of feature scores threshold can be four FI scores above and below the medianFI score, where the optimal subset of features can be found. This range holds good in all datasets and is exhibited in Table 4.3 – Table 4.6. Hence, the search space of feature subsets is reduced by 30% compared to performing exhaustive search on the feature subsets.

4.6 Evaluation of Performance Metrics of the Classifiers

Table 4.7 provides the performance summary of the classifiers on the heart datasets. CatBoost classifier with feature selection by FI type PVC of CatBoost scores the top precision, accuracy, recall, and f1-score on Cleve dataset and SA heart dataset. Hard MVE classifier modeled on PVC type selected features, shows the highest recall, accuracy, and f1-score on the Stat Heart Dataset and Hard MVE modeled on gain type selected features secures the highest values in recall, f1-score, precision and accuracy on Z-Ali heart dataset in comparison to other classifiers.

Table 4.7 Summary of Performance of the Classifiers with feature selection by FI Score

Classifier and FI type	**Cleveland HD Dataset**				**Statlog HD Dataset**			
	precision	**recall**	**f1-score**	**accuracy**	**precision**	**recall**	**f1-score**	**accuracy**
CatBoost-PVC	96%	85%	90%	**91.8%**	88%	82%	85%	85.2%
XGB-gain	95%	77%	85%	88.5%	86%	86%	86%	85.2%
MVE-PVC	92%	85%	88%	90. 2%	89%	**86%**	87%	**87.0%**
MVE-gain	95%	77%	85%	88.5%	89%	86%	87%	87.0%
Classifier and FI type	**South African HD Dataset**				**Z-Alizadeh Sani HD Dataset**			
	precision	**recall**	**f1-score**	**accuracy**	**precision**	**recall**	**f1-score**	**accuracy**
CatBoost-PVC	80%	57%	67%	**78.5%**	93%	92%	93%	93.0%
XGB-gain	75%	47%	58%	72.0%	98%	94%	96%	95.40%
MVE-PVC	73%	54%	62%	75.3%	98%	94%	96%	95.40%
MVE-gain	76%	50%	60%	73.1%	96%	98%	97%	**96.15%**

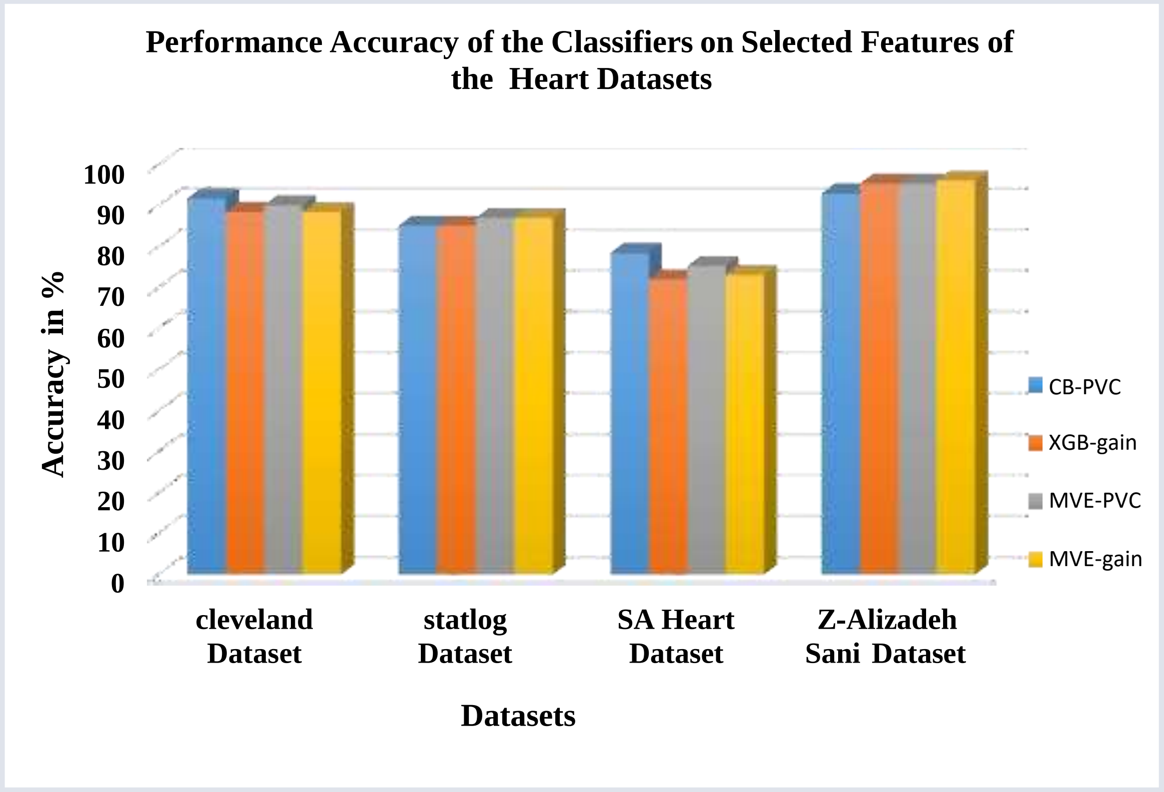

Figure 4.10 Performance Accuracy of the Classifiers on Selected Features of Heart Datasets

Figure 4.10 provides the top accuracy scores of XGBoost, CatB, and Hard MVE classifiers modeled on subsets of selected features. It is evident that CatB outperforms other classifiers on both Cleve and SA heart datasets and the Hard MVE classifier gives higher accuracy than others on the Stat and Z-Ali dataset.

4.7 Comparison of the Proposed and Existing Feature Selection Methods

Results from other feature selection methods are contrasted with those from the proposed Feature selection by FI Ranking approach. The datasets utilized to evaluate the proposed method each undergo separate feature selection processes. Table 4.8 contrasts the proposed feature selection with similar approaches from the literature, across all four datasets.

Table 4.8 Comparison of Existing work and Proposed work on Heart Datasets

Heart Dataset	Methods Adopted to select Features	Classifier	Accuracy
Cleveland Heart Disease Dataset	CFS Subset Evaluation+ Best-First-Search (Shan Xu et al., 2017)	Random Forest	92%
	Accumulating features ranks by Information gain, reliefF, gain ratio, oneR, and symmetric uncertainity (Ottom & Alshorman, 2019)	K Nearest Neighbors	79%
	Average of the FI done by Relief-F, Gain Ratio, Information Gain, and Chi-Square (Kolukisa et al., 2018)	Naive Bayes	86%
	Proposed Method	**CatBoost**	**92%**
Statlog Heart Disease Dataset	F-score (Divya & Agarwal, 2014)	Least Square Twin Support Vector Machine	86%
	RF (Hera et al., 2022)	Multi-Tier Ensemble (MTE)	84%
	Infogain, Correlation and ReliefF (Zahangir et al., 2019)	Random forest	83%
	Proposed Method	**Majority Vote Ensemble**	**87%**
SA heart Dataset	Infogain, Correlationand ReliefF (Zahangir et al.,2019)	Random forest	77%
	Mann Whitney-Wilcoxon Test and Logistic Regression (Babic et al., 2017)	Decision Tree	74%

Heart Dataset	Methods Adopted to select Features	Classifier	Accuracy
	Average of the FI done by Relief-F, Gain Ratio, Information Gain, and Chi-Square (Kolukisa et al., 2018)	SVM	75%
	Proposed Method	**CatBoost**	**78.5%**
Z-Alizadeh Sani Heart Dataset	Linear Discriminant Analysis (Kolukisa et al., 2018)	SVM	93%
	t-test and Principal Component Analysis (PCA) (Cüvitoğlu & Işik, 2018)	Artificial Neural Networks (ANN)	85%
	Genetic Algorithm (Arabasadi et al., 2017)	Neural Networks	94%
	Chi square (Dahal & Gautam, 2020)	SVM	89%
	Proposed Method	**Majority Vote Ensemble**	**96.15%**

4.8 CHAPTER SUMMARY

Feature Selection by Importance Scores of features based on ‘gain’ type of XGBoost and ‘Prediction Values Change (PVC)’ type of CatBoost were investigated. Feature subsets were formed by taking the importance score of each feature as the threshold. XGBoost, CatB, and Hard MVE classifiers were modeled on these feature subsets, and the subset which gave the best accuracy score was chosen. CatBoost classifiermodeled on Features selected by PVC type secured the top accuracy on Cleveland HD dataset and SA heart dataset. Hard MVE classifier modeled on PVC type selected features of the Statlog HD Dataset and Hard MVE modeled on gain type selected features of Z- Alizadeh Sani HD dataset exhibited high performance compared to other classifiers.

As an alternative to the exhaustive search done in related works, this work demonstrates that, across the subsets of features formed based on the FI Score by forward selection of top ranking features, it is ideal to search only the threshold range of four FI scores above and below the median FI score for the best subset of features. This search space identified would reduce the searches to be done when compared to exhaustive search of feature subsets.

Thus, feature selection by Importance Scores of Features of Gradient Boosting Algorithms is proposed with a significant reduction in the search space of the subsets of features for heart datasets.