Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5, 1454-1471
2024
Publisher: Learning Gate
DOI: 10.55214/25768484.v8i5.1848

Impact of feature selection techniques on machine learning and deep learning techniques for cardiovascular disease prediction-an analysis

Lijetha. C. Jaffrin[1*], J. Visumathi[2]
[1,2]Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology; lijethacjaffrin@veltech.edu.in (L.C.J.) drvisumathij@veltech.edu.in (J.V.).

Abstract: Cardio vascular disease is one of the life-threatening diseases which affects individuals worldwide. Early diagnosis may allow for the prevention or mitigation of cardiovascular diseases, which may minor mortality rates. A feasible Deep Learning and Machine Learning algorithms are used to find risk variables. Machine Learning and Deep Learning system anticipates heart diseases early on and reduce death rates from clinical data. To detect heart diseases or determine the patient's severity level, numerous research studies recently used various machine learning techniques. The volume of internationally recognised medical data sets is growing in terms of both qualities and records. This paper delivers brief outline of various feature extraction methods such as LASSO, RELIEF, RFE, MR-MR and RELIEF on deep learning and machine learning techniques for diagnosing cardiac disease. The performance metrics taken into consideration are Accuracy, Precision, Recall, F1score and the error measures are least Mean Squared Error and Mean Absolute Error. The feature selection methods with more features selected outpaced other approaches. Finally, crucial findings from the evaluated studies are outlined.

Keywords: *Deep learning, Features selection, Heart disease, Intelligent system, Machine learning.*

1. Introduction

The largest reason of death worldwide is heart disease, usually referred to as cardiovascular disease (CVD). According to the World Heart Federation's 2023 World Heart Report, nearly one-third of all fatalities worldwide were caused by cardiovascular diseases. Statistics from World Health Organisation indicates that more than 23.6 million people could die largely from heart disease and stroke in 2030 [1]. Stress, alcohol, smoking, unhealthy eating habits, sedentary lifestyles, and other factors like diabetes or high blood pressure are few causes of CVD [2]. Due to several factors, including accuracy and execution time, current heart disease diagnosis procedures are not very effective in detecting the disease in its early stages. As a result, researchers are working to develop an effective method for detecting the condition. When contemporary equipment and qualified medical personnel are not available, diagnosing and treating heart disease is really challenging. People's lives can be saved by earlier diagnosis and appropriate care. The primary goal is to provide succinct outline of feature selection methods on current Deep Learning and Machine Learning to foreknow heart disease.

A. Background Knowledge on ML and DL

Artificial Intelligence (AI) is a technology that makes it possible for computers and other devices to simulate decision making, human learning, conception, problem solving. These tasks comprise pattern recognition, experience-based learning, problem-solving, comprehension of natural language and decision-making. To analyse enormous volume of medical data and find patterns, risk factors, and early indicators of wide range of diseases, artificial intelligence (AI) is revolutionizing diagnosis and treatment. Enhancing early diagnosis, tailoring treatment regimens, and improving patient outcomes

* Correspondence: lijethacjaffrin@veltech.edu.in

are possible benefits of this. It is used often in the medical field to identify different diseases, including heart disease, cancer and diabetes. The idiom "artificial intelligence" (AI) includes analytical algorithms that repeatedly acquire from data, permitting computers to discover unobserved intuitions. When conventional statistical methods fall short, there are set of operations that include reinforcement learning, deep learning, cognitive learning and machine learning-based methods for combining and analysing complex healthcare and biological data [3]. The below fig.1 describes the features of Deep Learning, Artificial Intelligence and Machine Learning.

Machine learning (ML) is a subdivision of artificial intelligence that practices mathematical models for aiding computer learn without being overtly trained. Machine learning uses algorithms to discover patterns in data. A data model that forecasts the future is also constructed using these patterns. With the aim of delivering accurate diagnostics, risk assessment, and personalised treatments, machine learning algorithms are employed in prospective clinical trials [3]. The scope of machine learning (ML) is to impart computer to utilize its prior experience to crack problems. The notion of using machine learning to solve issues more quickly than humans in many fields. Perceptions and connections between data that are hidden to human eye can be found by processing and analysing very large quantity of data. Its smart behaviour is utilised in various algorithms that permit computer to make relevant prediction by conceptualizing information from experience [4].

Deep learning [4] is the division of machine learning resembling the function and structure of human brain. By automatically extracting complex patterns from large and complicated datasets, deep learning has revealed great potential in improving the diagnosis of cardiac disorders. Multiple layers in deep learning models process data through nonlinear transformations to obtain high-level characteristics. These models have revealed great accomplishment in various fields, such as picture recognition, speech recognition and natural language processing and, healthcare for applications like disease prediction, medical image analysis, and personalized treatment. Specifically, deep learning is the collection of neural data driven techniques built using automatic feature engineering procedures; its remarkable high-performance and accuracy stems from the ability to automatically learn features from inputs.

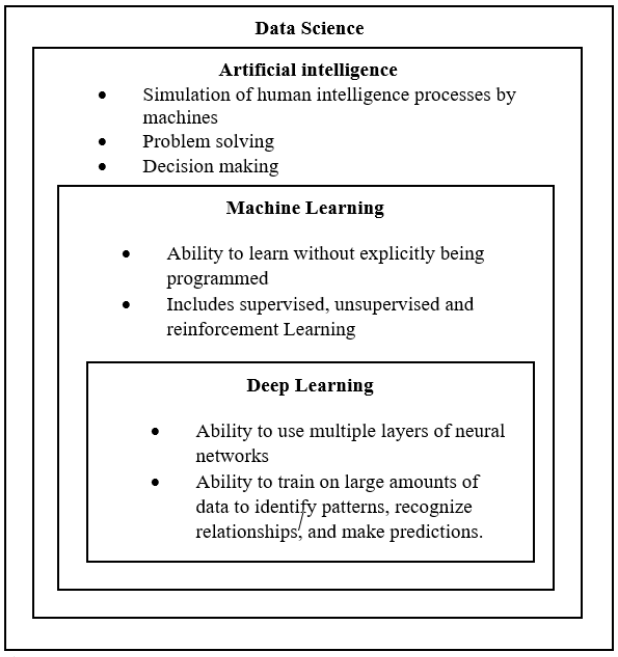

Figure 1.
What is AI, ML and DL?

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1445-1471, 2024
DOI: 10.55214/25768484.v8i5.1848

This paper encompasses the subsequent significant contributions:

- The aim was to look at the contribution of various feature selection methods, such as LASSO, RELIEF, MR-MR, RFE, and PCA impact Deep Learning and Machine Learning algorithms for prognosis of cardio vascular disease.
- Different feature selection approaches were used in the next step, including Random Forest, Support Vector Machine, Logistic Regression, Neural Network, Naïve Bayes and K-Nearest Neighbour. The outcomes were compared using Accuracy, Precision, Recall, F1score, MSE, and MAE assessment criteria.
- The primary and most noteworthy outcome of the current investigation is thorough comparison of several feature selection strategies on Deep learning and machine learning algorithms for forecasting cardiac illnesses.

2. Literature Review

The authors [1] proposed an efficient decision support system incorporating Improved Weighted Random Forest (IWRF) is used to forecast cardiac disease, supervised Infinite feature selection (Inf-FSs) to identify the most important features, and Bayesian optimization to adjust the new hyperparameters. It was supposed that coalescing these approaches would improve efficacy of current methodologies for predicting CVD using medical data sets. An Improved Weighted Random Forest was established to resolve class imbalance. The authors [5] utilised rank values in medical references and the most appropriate features were extracted using Least Absolute Shrinkage and Selection Operator and Relief feature selection. Overfitting and underfitting issues with machine learning were also solved with this. By mingling traditional classifiers with boosting and bagging methods, novel hybrid classifiers like, Gradient Boosting Method, AdaBoost Method, Decision Tree Bagging Method, K-Nearest Neighbours Bagging Method and Random Forest Bagging Method were created to increase testing rate and shorten implementation time [5].

The authors in [6] attempted to resolve the issue of feature selection by utilising pre-processing methods and four feature selection algorithms, including LASSO, Relief, LLBFS and MRMR for the proper subdivision of attributes. These features were employed for successful classifier training and testing. To choose features, the fast-conditional mutual information feature selection algorithm was secondly proposed. The authors [7] suggested a model for predicting diabetes using combined machine learning technique. The Support Vector Machine and Artificial Neural Network models brought up conceptual basis. The output of these models served as input membership function of fuzzy model and fuzzy logic concluded whether diabetes diagnosis was positive or negative. The fused models were saved for later use in a cloud storage system. A hybrid intelligent machine learning-based prediction technique [8] to forecast cardiac sickness used different algorithms, including K-nearest neighbour, Random Forest, Decision Tree, Linear Discriminant Analysis, Support Vector Machine and Gradient Boosting Classifier as well as feature selection approach Sequential Feature Selection. By implementing K-fold cross-validation method, Sequential Feature Selection approach can reduce calculation time while increasing the classification accuracy of classifiers.

The authors [9] presented an ensemble model and majority voting method to improve the accuracy of prediction of heart disease. Additionally, feature selection based genetic algorithm and pre-processing method were proposed to improve prediction accuracy. For predicting CHD using an improved LightGBM classifier, HY_OptGBM was proposed as prediction model [10]. The hyperparameters of the LightGBM model were altered to improve the classifier's performance. The model was trained utilising altered hyperparameters, and its loss function was enhanced as well. The most cutting-edge hyperparameter optimisation framework (OPTUNA) was used to optimise the prediction model's

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1445-1471, 2024
DOI: 10.55214/25768484.v8i5.1848

hyperparameters. For the purpose of predicting cardiovascular disorders utilising 12 lead-based ECG images, a new, lightweight deep learning CNN architecture was suggested [11]. The suggested CNN model outperformed the cutting-edge low-scale Squeeze Net and AlexNet, with success rates.

An effective method for foreseeing risk of coronary heart disease was proposed [12]. It incorporated deep neural networks that were trained on well-organized training data. First, it demonstrated how Principal Component Analysis and Variational Autoencoder models enhance functionality of deep neural network. In second experiment, proposed approach was contrasted with Adaptive Boosting, Support Vector Machine, Decision Tree Naive Bayes, K-Nearest Neighbour, and Random Forest. Experimental outcomes demonstrated that suggested approach beat traditional machine learning methods by providing better accuracy. Using the UCI heart illness dataset, a novel method [13] for Ischemic Heart illness Squirrel Search Optimisation feature selection technique was proposed. In the study, characteristics like "Cp," "restecg," "oldpeak," "Ca," and "thal" were identified for prognosis of heart disease. The suggested ischemic heart disease squirrel search optimization model and random forest classifier provided enhanced accuracy for heart disease prediction.

A fused Random Forest with Linear Model was presented [14]. Its major goal was to improve the predictability of cardiac disease. HRFLM approach, in contrast, had been an advantage of every feature without any feature selection limitations and had higher ability to predict heart disease than existing methods. Recursion Enhanced Random Forest with Improved Linear Model was suggested [15] for predicting heart disease and evaluating its accuracy. It was additionally suggested to design artificial neural network with feature selection and backpropagation learning for categorising cardiovascular illness. The authors in [16] suggested to use improved SALP swarm optimisation and an adaptive neuro-fuzzy inference system framework for detection of cardiac illness in order to improve prediction accuracy. Utilising Levy flight algorithm, the suggested model enhanced search capability. Consistent learning in adaptive neuro-fuzzy inference system relied on gradient-based learning and was a prone for getting jammed in local minima. modified SALP swarm optimization- adaptive neuro-fuzzy inference system performed well in terms of disease prediction, as demonstrated by a simulation and analysis.

OCI-DBN, a unique method that did not overfit or underfit, was suggested [17] for categorization and improvement in heart disease prognosis. In the suggested OCI-DBN strategy, Ruzzo-Tompa performed best. By determining each hidden layer in DBN using SGA, proposed OCI-DBN technique resolved network optimisation and setup issue. In this study [18], the properties of big data and machine learning for foreseeing the risk of CVD were investigated using Korean National Health Insurance Service-National Health Sample Cohort data analysis. More particularly, accuracy of different machine learning (ML) techniques such as LightGBM, deep neural networks, logistic regression, and random forests in predicting 2-years and 10-years risk of CVD, including heart failure, atrial fibrillation, strokes and coronary artery disease was evaluated. A sophisticated learning system Random Search Algorithm based Random Forest [19] was suggested for automatic recognition of heart failure. This model for heart failure diagnosis was put forth and developed. In proposed diagnostic system, features were selected using random search algorithm, and heart failure was foreseen using random forest model. Grid search was used to optimise suggested diagnostic system. In order to effectively predict Heart Failure, an expert system [20] comprising two support vector machine models was presented. Linear and L1 regularised described first SVM model. L2 regularised was second SVM model. It served as a predictive modelling tool. In order to optimise two models, hybrid grid search algorithm was suggested that instantaneously optimised two models. Matthew's correlation coefficient, Accuracy, specificity, area under the curve, receiver operating characteristic charts, and sensitivity [20] were six evaluation metrics that evaluate effectiveness of recommended approach.

To increase diagnosis efficiency of heart illness, a smart hybrid system [21] called χ^2 -DNN had been created. This study examined the consequence of neural network depth on accuracy. Deep neural networks are generally achieved better on tiny datasets. However, DNN with multiple hidden layers performed better than ANN on heart disease dataset and unnecessary characteristics are removed. A two-level stack model [22] was created, with level-1 serving as base level and level-2 as meta level. The

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1445-1471, 2024
DOI: 10.55214/25768484.v8i5.1848

input for meta-level classifiers was preferred from predictions of base-level classifiers. To identify lowest correlation classifier, maximum information coefficient and Pearson correlation coefficient were first calculated. Then, best combining classifiers that produced best result were found using enumeration approach. This study [23] analyses the 299 hospitalised patients with heart failure who survived and used nine classification models to predict patient survival: Decision Tree, Stochastic Gradient classifier, Adaptive boosting classifier, Random Forest, Gradient Boosting classifier, Logistic Regression, Gaussian Naive Bayes classifier Support Vector Machine and Extra Tree Classifier. The class imbalance issue was resolved by Synthetic Minority Oversampling Technique.

A highly accurate automated method [24] for predicting sudden cardiac death utilising quantifiable arrhythmic signals was described. Three repolarization interval ratios as well as two conduction-repolarization indicators were included in arrhythmic parameters set. Then, all determined markers were applied to automatic classification of normal and sudden cardiac death risk groups using machine learning classifiers such k-nearest neighbour, random forest, Naive Bayes, support vector machine and decision tree. The current findings demonstrated that both automatic classifier and integrated sudden cardiac death index could predict sudden cardiac death up to 30 minutes earlier. For ensemble voting algorithms to predict cardiac disease [25], novel grouping of machine learning classifiers is presented. Individual classifiers are first empirically evaluated for their ability to forecast heart disease, and then are chosen for ensemble voting systems based on their performance. Four heart disease datasets are specifically used for performance evaluation of six single classifiers and five ensemble voting schemes. The proposed ensemble approach has produced improved outcomes than other schemes. An effective and highly accurate prediction of cardiovascular disorders was made possible by Machine Learning based Cardiovascular Disease Diagnosis architecture [26]. The framework initially addressed missing values in dataset using mean replacement technique and data imbalance using SMOTE. The comparison analysis concluded that MaLCaDD predictions were more accurate. The authors [27] [28] evaluated and provided an overview of the general prediction power of ML algorithms in cardiovascular disorders. SVM might perform better than other algorithms, even though there were lack of studies in meta-analytics for both heart failure and cardiac arrhythmias.

The overall objective [29] was to predict precisely the presence of heart disease. Simulation-based experimentations were shown using six methodologies named Naive Bayes Classifier, K-Nearest Neighbour, Logistic Regression, Decision Tree Classifier, Support Vector Machine and Random Forest. It was concluded that Random Forest produced more accuracy than other techniques. This study [30] sought to find several supervised machine-learning algorithms for predicting cardiac disease. The utility of Machine Learning techniques for heart illness prediction were tested on heart disease dataset, and it was revealed that three classification algorithms, k-nearest neighbour, random forests and decision tree algorithms performed flawlessly, achieving an excellent performance. Authors [31] identified the most effective ML algorithm Naive Bayes and Decision Tree algorithms for heart disease identification. The study's findings presented that Random Forest algorithm was effective algorithm for predicting heart disease, with enhanced accuracy and produced better results. For forecasting the survival of heart patients, the authors [32] suggested machine learning-based technique to predict cardiac disease. They used eight classification models, including Linear Discriminant Analysis, Decision Tree, Adaptive Boosting, Extra Tree, Ridge Classifiers, Logistic Regression, Random Forest, and Light Gradient Boosting Machine. SMOTE was utilised to solve class inequality issue. The authors [33] used machine learning techniques like support vector machine, random forest, K-nearest neighbour, and deep learning models like gated-recurrent unit neural networks and long short-term memory to deploy on UCI Heart Disease dataset. By integrating various classifiers, ensemble voting-based models had been used to enhance the accuracy of weak algorithms.

To forecast the presence of cardiac disease in an individual, authors [34] used supervised machine learning methods. The heart illness dataset was subjected to Naive Bayes, K-Nearest Neighbours and Random Forest in order to determine prediction accuracy. Accuracy of prediction models was improved by handling null values on specific column and imputing mean values for that column with information

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1445-1471, 2024
DOI: 10.55214/25768484.v8i5.1848

gain feature selection method. Exploratory data analysis was conducted on heart dataset by authors [35]. Utilising the Python Seaborn modules and Spyder IDE, exploratory data analysis had been done. Distribution plots, Matrix plots, Categorical plots, Regression plots, Grid plots and advanced plots have been created for heart dataset. The variables prevalentHyp, sysBP, diaBP, heartRate, totChol, and glucose were chosen as essential subcategory of variables from dataset's 16 variables to predict progress of heart disease in individuals in different age groups.

3. Methodology

The suggested method began with the collection of five open-source heart datasets Cleveland, statlog, Hungarian, Switzerland and Long Beach. As standard bench mark datasets were combined and used, no pre-processing was required. Following the data collection, five feature selection strategies such as LASSO, RELIEF, MR-MR, RFE, PCA were used. The optimal subset was then chosen, and six machine-learning strategies were used. The data is subjected to following machine learning algorithms: random forest, K-Nearest Neighbour, neural network, Support Vector Machine, Naïve Bayes and Logistic Regression. The performance of algorithm and feature selection was then assessed using a variety of evaluation criteria. Information of 1190 cardiac patients is included in the combined heart disease dataset. Heart disease was classified as stage 1, 2, 3, 4, or no heart disease at the time of diagnosis for each patient. The dataset contains 1190 data records. Following are the dataset's features: exercise angina, sex, fasting blood sugar, resting ecg resting bps, chest pain type, max heart rate, age, cholesterol, target, ST slope, oldpeak.

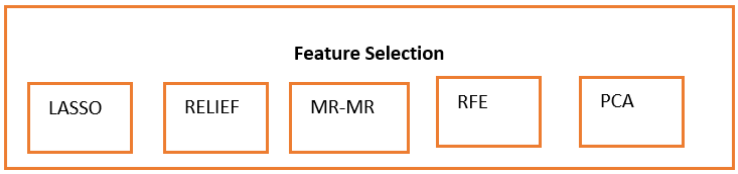

Figure 2.
Different feature selection methods.

3.1. Feature Selection Methods

Feature selection [36] means choosing subclass of input features from data to improve performance of prediction model and minimize noise. By selecting just pertinent data and eliminating noise from the data, it is a technique for lowering the input variable of your model. The objective is to optimize model's performance with least amount of data while streamlining and clarifying the model. The process of automatically selecting pertinent characteristics according to the kind of problem being solved is known as feature selection. To achieve this, choose which significant characteristics to include or remove without altering them. Reducing the size of the input dataset and removing unnecessary noise from the data are two benefits of the technique. The above Fig. 2 shows different feature selection techniques employed in this work and they are described below.

3.1.1. Least Absolute Shrinkage and Selection Operator (LASSO)

Least Absolute Shrinkage and Selection Operator is a statistical technique used in statistics and machine learning [37]. This particular kind of regression analysis carries out two crucial tasks: Variable selection is the process of determining which characteristics or factors are most crucial for predicting an outcome variable. Regularization: It lowers the model's variance, which lessens the likelihood of overfitting and enhances the model's applicability to hypothetical data. LASSO augments the conventional least squares objective function with a penalty term to accomplish these ends. The absolute value of the regression coefficients serves as the basis for this penalty term. Certain coefficients can be

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1445-1471, 2024
DOI: 10.55214/25768484.v8i5.1848

efficiently set to exactly zero via LASSO by reducing the coefficients towards zero. This leads to variable selection since the matching variables are eliminated from the model.

An additional penalty term has been added to objective function by LASSO. The absolute value of the coefficients serves as the foundation for this penalty term. A coefficient's penalty increases with its absolute value. LASSO works to reduce the coefficients to zero by minimizing the combined objective function, which considers both the LASSO penalty and the squared errors. Coefficients may occasionally be driven all the way to zero, which would essentially eliminate the associated characteristics from the model. Better generalization performance on unobserved data results from this shrinking, which also lowers the model's variance. Furthermore, LASSO can enhance the model's interpretability by limiting the selection to the most crucial elements.

The LASSO optimization problem can be expressed as

$$\text{Min } [(1/M) * \Sigma(c_i - (\beta_0 + \Sigma(\beta_k * a_k)))^2 + \lambda * \Sigma|\beta_k|] \quad [1]$$

where M is the count of data points, c_i is the real value of i-th data point for target variable, β_0 is the intercept of the model, β_k is the coefficient for k-th feature, a_k is the value of k-th feature for i-th data point, Σ is the summation over all features (k) and λ is the Tuning parameter.

3.1.2. RELIEF

Relief feature selection [38] is a filter-based technique that extracts the most appropriate features from dataset. It concentrates on choosing characteristics that are most important for distinguishing between examples that belong to various classes. Its purpose is to remove superfluous or unnecessary features, which results in faster training times because fewer features mean less computation, better interpretability because simpler models are easier to understand, and improved model performance because reduced complexity prevents overfitting and improves generalization.

Relief [39] depends on the idea of closest neighbours. It determines the closest neighbours from several classes (the most comparable points falling into various categories) for a given data point. Each feature is examined, and the values of the features at the original position and their closest neighbours are compared. It distinguishes between two situations: A "hit" is an unwanted result that occurs when nearest neighbour of different class has same feature value. Miss: A "miss" (desirable) occurs when the feature value of nearest neighbour (of different class) differs. Relief adjusts a score according to the quantity of "hits" and "misses" experienced for every feature. High marks are awarded to features that have a lot of "misses" (differentiate classes) and few "hits" (constant within class).

Let D represent the dataset consisting of n occurrences and p features.

For each feature, the relevance score is stored in a weight vector W of zeros (length p), and T is the number of iterations (usually high for better estimation). For every cycle (t = 1 to T), as follows:

1. One instance R_a should be chosen at random from D.
2. Determine which R_a feature values correspond to the nearest hit (H) and nearest miss (M) (same class for H, different class for M).
3. Modify each feature b's weight (from 1 to p):

$$W[b] = W[b] - \text{diff}(R_a[b], H[b]) / (T * m) + \text{diff}(R_a[b], M[b]) / (T * m) \quad [2]$$

Now, diff(a, b) determines how feature values a and b differ from each other. The number of nearest neighbours (usually fixed) that are taken into consideration is m.

The weight vector (W), which shows the significance of each feature after T iterations, has greater values that correspond to better class separation.

3.1.3. Minimum Redundancy Maximum Relevance (MR-MR)

The Goal of Minimum Redundancy Maximum Relevance [40] is to categorize subclass of features that are less redundant with one another and extremely significant to the target variable.

Relevance: MR-MR measures the significance of a feature X_i's relationship to the target variable Y by using mutual information. Two variables are dependent on each other, and mutual information

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1445-1471, 2024
DOI: 10.55214/25768484.v8i5.1848

quantifies this. The association between feature and goal variable is stronger when there is a higher mutual information.

Relevance $(X_i, Y) = I(X_i, Y)$

Redundancy: MR-MR measures how much information a feature X_i shares with another feature X_j. The goal is to prevent overfitting in machine learning models by avoiding using features that represent the same information. This instance likewise uses mutual information, but it does so to gauge how dependent two sets of features are on one another. When features have high redundancy, they offer overlapping information [41].

Redundancy $(X_i, X_j) = I(X_i, X_j)$

Average Redundancy: The average redundancy between feature X_i and every feature in the chosen set S is the redundancy term

$$\frac{1}{|S|} \sum_{X_j \in S} \text{Redundancy}(X_i, X_j)$$

MRMR Score combines redundancy and relevance to choose features that are least redundant with one another and extremely relevant to the target variable. The following formula can be used to express the MR-MR (Minimum Redundancy Maximum Relevance) feature selection Score mathematically:

$$\text{MRMR Score}(X_i) = \text{Relevance}(X_i, Y) - \frac{1}{|S|} \sum_{X_j \in S} \text{Redundancy}(X_i, X_j) \quad [3]$$

where:

- MRMR Score (X_i) is MR-MR score for feature X_i.
- Relevance (X_i, Y) is relevance of feature X_i with respect to the target variable Y.
- S is the subset of already chosen features.
- |S| is count of features in subset S.
- Redundancy (X_i, X_j) is the redundancy among feature Xi and each feature X_j in selected set S.

3.1.4. Recursive Feature Elimination (RFE)

Recursive Feature Elimination method [41] helps to identify characteristics which are most crucial to include in a model. Depending on the model's coefficients, it operates by iteratively eliminating least significant features until predetermined number of features is reached. Recursively evaluating progressively smaller subsets of features is how RFE selects features. Based on the feature importance or coefficients supplied by the model, it trains a model in every iteration and eradicates least significant feature. The importance of each feature is computed using Linear model-specific criteria as follows: Use absolute values of coefficients.

A linear model

$$y = q_0 + q_1 x_1 + \ldots + q_p x_p, \quad [4]$$

The importance of feature j is $|q_j|$

With RFE, redundant or unnecessary features are removed, which effectively reduces model complexity and improves generalization.

3.1.5. Principle Component Analysis (PCA)

To extract the most variance from a dataset, principle Component Analysis (PCA) [40] [41] divides the dataset into a set of orthogonal components, or principle components. PCA can be used to determine the most significant features by examining the principle components, even though it is not a feature selection technique per se. In order to recognize the original features which contribute the most to variance in the dataset, one must interpret the principle components when using PCA for feature selection. Despite the fact that PCA generates new composite features (principal components) instead of choosing original features, it can still influence feature selection by evaluating how important a feature is within these components. In order to choose features using PCA, the loadings of the principle components are examined. The loadings show how each original feature contributed to a major

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1445-1471, 2024
DOI: 10.55214/25768484.v8i5.1848

component. They are given by the matrix of eigenvectors: L = V, where each element Lij in the loading matrix represents the contribution of the j^{th} feature to the i^{th} principal component.

4. Results and Discussion

4.1. Results Computed with LASSO Feature Selection

8 features have been chosen with LASSO from initial features set of 12 features from Combined dataset. Table I depicts the performance of LASSO feature selection with six implemented ML and Neural network classifiers in terms of six vital efficiency measuring metrics. Figure 3-8 signifies the six outcomes Accuracy, Precision, Recall, F1score, MSE, MAE [42] [43] in graphical format for Random Forest, Logistic Regression, Support Vector Machine, Naive Bayes, K-Nearest Neighbour classifiers respectively using LASSO feature selection.

Table 1.
Results with lasso feature selection method.

LASSO feature extraction						
Model	**Accuracy**	**Precision**	**Recall**	**F1score**	**MSE**	**MAE**
Logistic regression	86.13	87.12	87.78	87.45	13.86	13.86
KNN	88.65	87.14	93.12	90.03	11.34	11.34
Random forest	95.37	94.77	96.94	95.84	4.62	4.62
SVM	89.07	86.71	94.65	90.51	10.9	10.9
Naive bayes	85.71	86.46	87.78	87.12	14.28	14.28
Neural network	95.92	96.35	97.45	95.21	4.02	4.02

4.2. Results Computed with RELIEF Feature Selection

Using the feature selection algorithm RELIEF, 10 features were obtained from the original set of 12 features in Combined dataset. Table II reveals the outcome of RELIEF feature selection using six different ML classifiers, broken down into six essential efficiency measurement criteria. The six outcomes Accuracy, Precision, Recall, F1score, MSE, and MAE [42] [43] are shown graphically for the classifiers of Random Forest, Support Vector Machine, Naive Bayes, Logistic Regression, K-Nearest Neighbour respectively, in Figure 3-8.

Table 2.
Results with relief feature selection method.

RELIEF feature extraction						
Model	**Accuracy**	**Precision**	**Recall**	**F1score**	**MSE**	**MAE**
Logistic regression	77.31	81.81	75.57	78.57	22.68	22.68
KNN	78.15	82.64	76.33	79.36	21.84	21.84
Random forest	81.09	86.44	77.86	81.92	18.9	18.9
SVM	77.31	81.81	75.57	78.57	22.68	22.68
Naive bayes	78.99	84.03	76.33	80	21.00	21.00
Neural network	82.56	88.47	79.86	83.92	14.5	14.5

4.3. Results Computed with PCA Feature Selection

Using the feature selection technique PCA, nine features were obtained from the original set of twelve features in the combined dataset. Table III illustrates the result of PCA feature selection using six different ML classifiers, broken down into six essential efficiency measurement criteria. The six outcomes such as Accuracy, Precision, Recall, F1score, MSE, and MAE [43] [44] for each of the following classifiers— Support Vector Machine, Naive Bayes, Random Forest, K-Nearest Neighbour, Logistic Regression, neural network are shown graphically in Figures 3-8.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1445-1471, 2024
DOI: 10.55214/25768484.v8i5.1848

Table 3.
Results with PCA feature selection method.

PCA feature extraction						
Model	**Accuracy**	**Precision**	**Recall**	**F1score**	**MSE**	**MAE**
Logistic Regression	84.87	85.18	87.78	86.46	15.12	15.12
KNN	82.77	84.09	84.73	84.41	17.22	17.22
Random Forest	90.33	89.70	93.12	91.38	9.66	9.66
SVM	85.29	85.29	88.54	86.89	14.70	14.70
Naive Bayes	83.61	84.32	86.25	85.28	16.38	16.38
Neural Network	91.37	90.45	95.62	90.27	8.46	8.46

4.4. Results Computed with RFE Feature Selection

After deploying feature selection technique RFE, ten features were derived from the 12 initial features from the combined dataset. Table IV shows the RFE feature selection performance in terms of six essential efficiency measurement parameters when six ML classifiers are employed. For the classifiers of neural network, Support Vector Machine, Naive Bayes, Random Forest, Logistic Regression, K-Nearest Neighbour, respectively, the six outcomes such as Recall, Accuracy, Precision, F1score, MSE, and MAE [44] [45] are shown graphically in Figure 3-8.

Table 4.
Results with RFE feature selection method.

RFE feature extraction						
Model	**Accuracy**	**Precision**	**Recall**	**F1score**	**MSE**	**MAE**
Logistic regression	85.71	87.59	86.25	86.92	14.28	14.28
KNN	79.83	78.23	87.78	82.73	20.16	20.16
Random forest	85.29	85.82	87.78	86.79	14.7	14.7
SVM	86.55	86.13	90.07	88.05	13.44	13.44
Naive bayes	83.19	86.4	82.44	84.37	16.8	16.8
Neural network	87.77	87.82	89.78	88.34	13.2	13.2

4.5. Results computed with MR-MR Feature selection

Out of the original features set of 12 features from the combined dataset, 11 features have been obtained utilizing the MR-MR feature selection technique. Table V illustrates the results of MR-MR feature selection method using six real-world deep learning and Machine Learning classifiers, broken down into six essential efficiency measures. The six outcomes such as Accuracy, Recall, Precision, F1score, MSE, and MAE [44] [45] for K-Nearest Neighbour, neural network, Support Vector Machine, Random Forest, Naive Bayes, Logistic Regression classifiers are shown graphically in Figure 3.

Table 5.
Results with Mr-Mr feature selection method.

MR-MR feature extraction						
Model	**Accuracy**	**Precision**	**Recall**	**F1score**	**MSE**	**MAE**
Logistic regression	80.25	80.43	84.73	82.52	19.74	19.74
KNN	78.15	80.62	79.38	80	21.84	21.84
Random forest	90.33	90.29	92.36	91.32	9.6	9.6
SVM	84.87	86.82	85.49	86.15	15.12	15.12
Naive bayes	79.41	81.53	80.91	81.22	20.58	20.58
Neural network	91.24	91.12	93.16	92.62	8.11	8.11

Figure.3 below shows the accuracy of deep learning and machine learning algorithms after each feature selection method. After deploying feature selection method LASSO, accuracy of deep learning and machine learning classifiers such as Logistic Regression, Random forest, Support Vector Machine, Naïve Bayes, K-Nearest Neighbour, and Neural Network are 86.13, 95.37, 89.07, 85.71, 88.65, 95.92 respectively. After deploying feature selection method RELIEF, the accuracy of machine learning classifiers such as K-Nearest Neighbour, Random forest, Support Vector Machine, Naïve Bayes, Logistic Regression and Neural Network are 78.15, 81.09, 77.31, 78.99, 77.31, 82.56 respectively. With PCA, the accuracy of various machine learning and deep learning classifiers such as Logistic Regression, Random forest, K-Nearest Neighbour, Naïve Bayes, Support Vector Machine, and Neural Network are 84.87, 90.33, 82.77, 85.29, 83.61, 91.37 respectively. Using RFE feature selection, the accuracy of Logistic Regression, Random forest, Support Vector Machine, Naïve Bayes, K-Nearest Neighbour, and Neural Network are 85.71, 85.29, 83.19, 86.55, 79.83, 87.77 respectively. With MR-MR feature selection, accuracy of K-Nearest Neighbour, Random forest, Logistic Regression, Support Vector Machine, Naïve Bayes and Neural Network are 78.15, 90.33, 80.25, 84.87, 79.41, 91.24 respectively. The Random forest algorithm and Neural network classifier implemented using LASSO feature selection technique achieved high performance compared with other algorithms. Random forest and Neural Network classifiers achieved an accuracy [46] of 95.37 and 95.92 respectively.

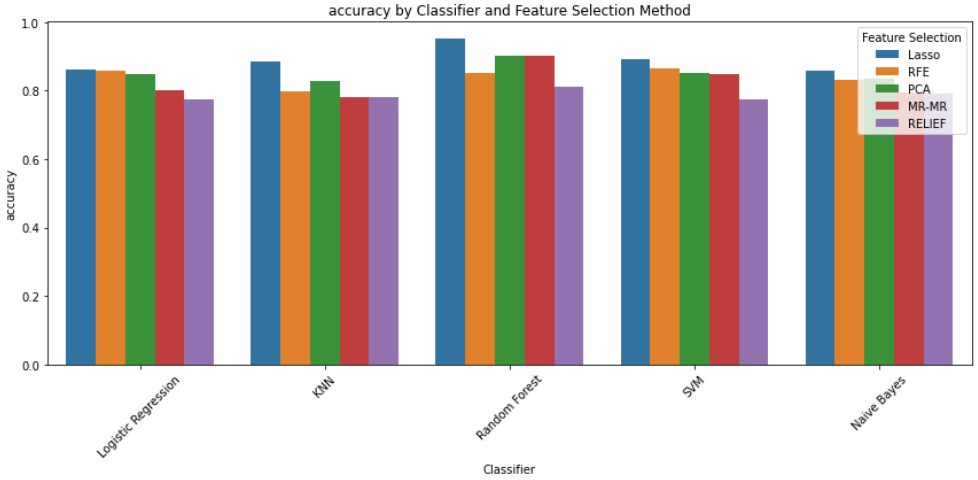

Figure 3.
Accuracy of different classifier with feature selection method.

Figure 4 below illustrates the precision of machine learning and deep learning algorithms after each feature selection method. After deployment of feature selection method LASSO, the precision of Random forest, Logistic Regression, K-Nearest Neighbour, Support Vector Machine, Neural Network and Naïve Bayes are 94.77, 87.12, 87.14, 86.71, 96.35, 86.46 respectively. With feature selection method RELIEF, the precision of Logistic Regression, Neural Network, K-Nearest Neighbour, Random forest, Support Vector Machine, Naïve Bayes and are 81.81, 88.47, 82.64, 86.44, 81.81, 84.03 respectively. The precision of Naïve Bayes, Logistic Regression, K-Nearest Neighbour, Random forest, Support Vector Machine, and Neural Network after PCA feature selection are 84.32, 85.18, 84.09, 89.70, 85.29, 90.45 respectively. The precision of Support Vector Machine, Logistic Regression, K-Nearest Neighbour, Random forest, Naïve Bayes and Neural Network after RFE feature selection are 86.13, 87.59, 78.23, 85.82, 86.4, 87.82 respectively. The precision of Random forest, Logistic Regression, K-Nearest Neighbour, Support Vector Machine, Naïve Bayes and Neural Network after MR-MR feature selection are 90.29, 80.43, 80.62, 86.82, 81.53, 91.12 respectively. The Random forest and neural network

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1445-1471, 2024
DOI: 10.55214/25768484.v8i5.1848

classifier implemented using LASSO and MR-MR feature selection techniques have achieved high precision compared with other algorithms.

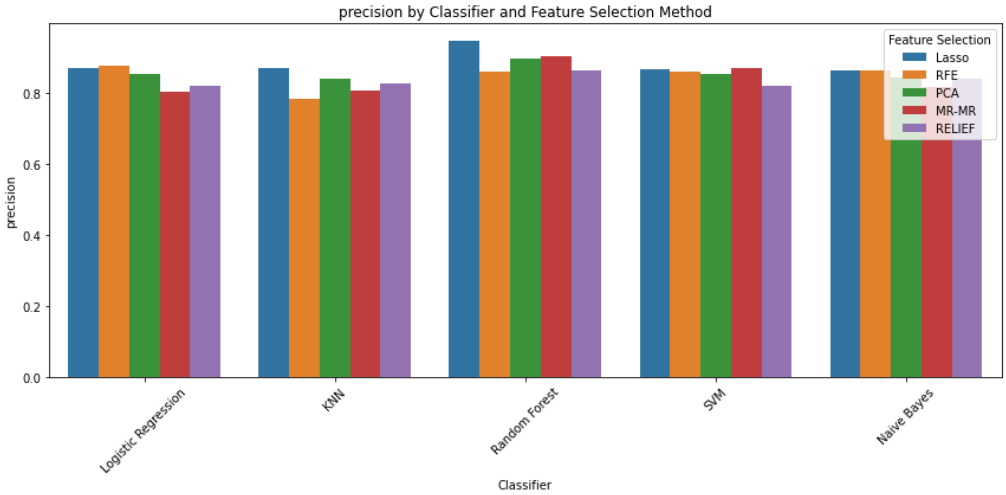

Figure 4.
Precision of different classifier with feature selection methods.

Figure 5 below shows the recall measure of machine learning and deep learning algorithms after each feature selection techniques. After deployment of feature selection method LASSO, the recall measure of Support Vector Machine, Logistic Regression, Naïve Bayes, K-Nearest Neighbour, Random forest, and Neural Network are 94.65, 87.78, 87.78, 93.12, 96.94, 97.45 respectively. With feature selection method RELIEF, the recall measure of Logistic Regression, K-Nearest Neighbour, Neural Network, Random forest, Support Vector Machine, Naïve Bayes are 75.57, 76.33, 79.86, 77.86, 75.57, 76.33, respectively. The recall measure of Logistic Regression, Support Vector Machine, K-Nearest Neighbour, Random forest, Support Vector Machine, Naïve Bayes and Neural Network after PCA feature selection are 87.78, 88.54, 84.73, 93.1286.25, 95.62 respectively. The recall measure of Neural Network, K-Nearest Neighbour, Random forest, Support Vector Machine, Naïve Bayes and Logistic Regression, after RFE feature selection are 89.78, 87.78, 87.78, 90.07, 82.44, 86.25 respectively. After the deployment of feature selection method MR-MR, the recall measure of Logistic Regression, K-Nearest Neighbour, Random forest, Support Vector Machine, Naïve Bayes and Neural Network are 84.73, 79.38, 92.36, 85.49, 80.91, 93.16 respectively. The Random forest and neural network classifier implemented using LASSO feature selection techniques have achieved high recall measure compared with other algorithms.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1445-1471, 2024
DOI: 10.55214/25768484.v8i5.1848

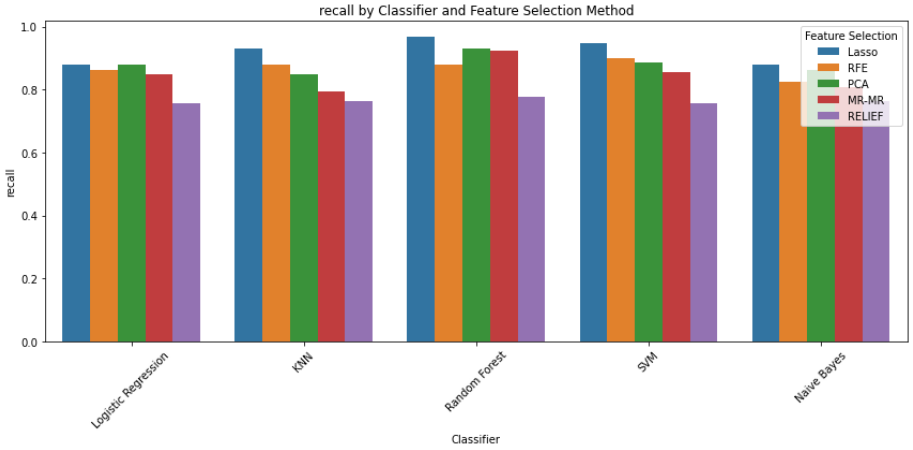

Figure 5.
Recall of different classifier with feature selection methods.

Figure.6 below displays the F1 score of deep learning and machine learning algorithms after each feature selection method. After application of feature selection method LASSO, the F1 score of K-Nearest Neighbour, Logistic Regression, Random forest, Support Vector Machine, Neural Network and Naïve Bayes are 90.03, 87.45, 95.84, 90.51, 95.21, 87.12 respectively. With feature selection method RELIEF, the F1 score measure of Logistic Regression, K-Nearest Neighbour, Support Vector Machine, Naïve Bayes, Random forest, and Neural Network are 78.57, 79.36, 78.57, 80, 81.92, 83.92 respectively. The F1 score measure of Logistic Regression, K-Nearest Neighbour, Neural Network, Support Vector Machine, Naïve Bayes and Random forest after PCA feature selection are 86.46, 84.41, 90.27 86.89, 85.28, 91.38 respectively. The F1 score of Logistic Regression, K-Nearest Neighbour, Random forest, Support Vector Machine, Naïve Bayes and Neural Network after RFE feature selection are 86.92, 82.73, 86.79, 88.05, 84.37, 88.34 respectively. After the deployment of feature selection method MR-MR, the F1 score measure of Logistic Regression, Neural Network Random forest, Support Vector Machine, Naïve Bayes and K-Nearest Neighbour, are 82.52, 92.62, 91.32, 86.15, 81.22, 80 respectively. The Random forest and neural network classifier implemented using LASSO feature selection techniques have achieved high F1 measure than other algorithms.

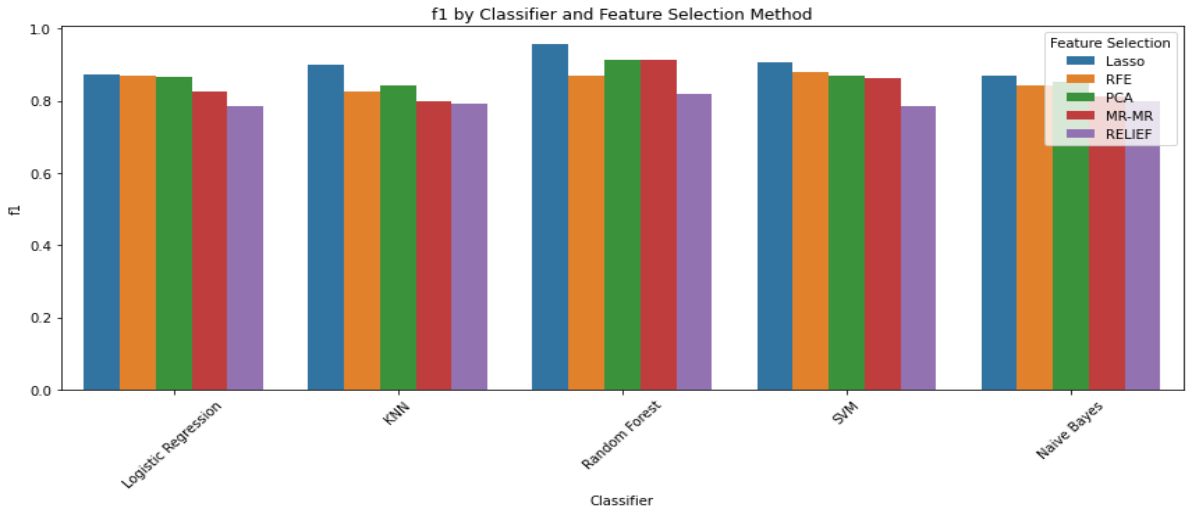

Figure 6.
Fl-measure of different classifier with feature selection methods.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1445-1471, 2024
DOI: 10.55214/25768484.v8i5.1848

Figure 7 below shows the Mean Squared Error (MSE) of deep learning algorithms and machine learning after each feature selection method. After implementing feature selection method LASSO, Random forest and Neural Network achieves high performance with least Mean Squared Error of 4.62 and 4.02 respectively compared to other algorithms.

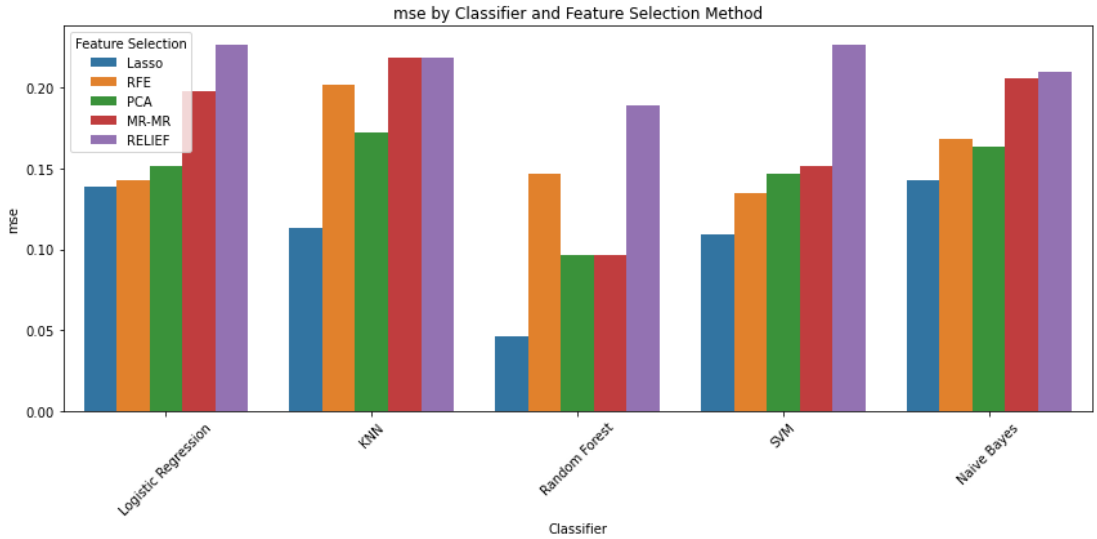

Figure 7.
Mean squared error of different classifier with feature selection methods.

Figure 8 below shows the Mean Absolute Error (MAE) of machine learning and Deep learning algorithms after each attribute selection method. After implementing feature selection method LASSO, Random forest and Neural Network achieves high performance with least Mean Absolute Error of 4.62 and 4.02 respectively compared to other algorithms.

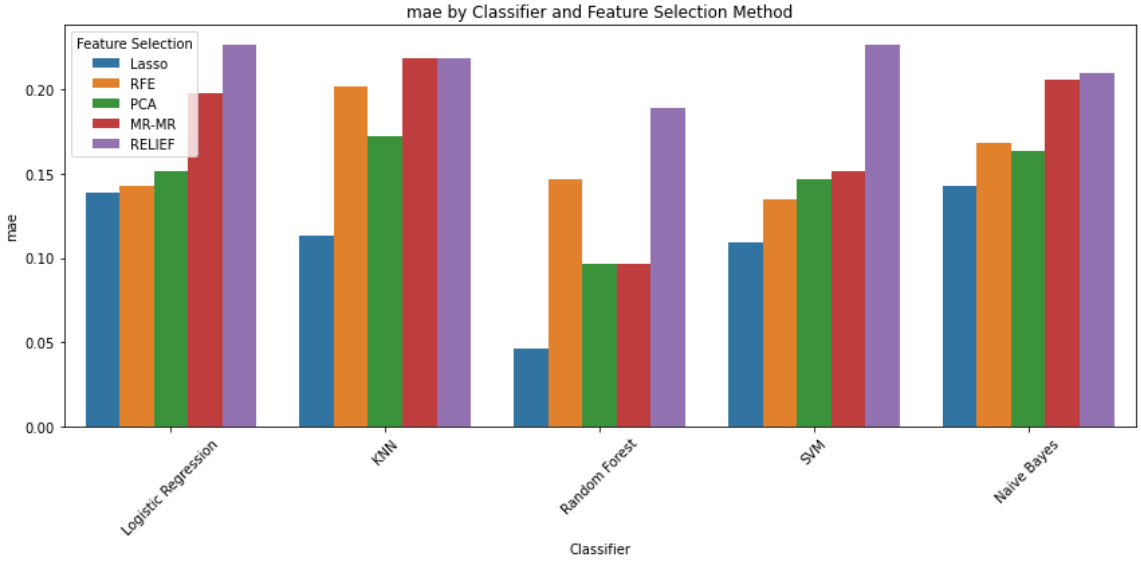

Figure 8.
Mean absolute error of different classifier with feature selection methods.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1445-1471, 2024
DOI: 10.55214/25768484.v8i5.1848

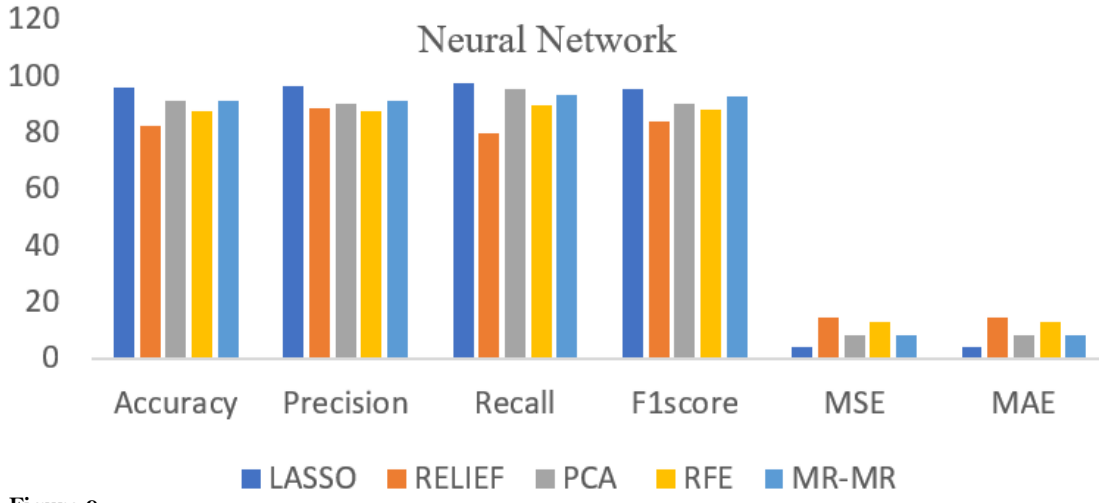

Figure 9.
Different feature selection on neural network.

Figure.9 above shows the Accuracy, Recall, Precision, F1score, Mean Squared Error and Mean Absolute Error of Neural Network after deploying different feature selection methods such as LASSO, RELIEF, MR-MR, RFE, and PCA. After implementing feature selection method LASSO, Neural Network technique achieved high performance with accuracy of 95.92%, Precision 96.35%, Recall 97.45%, F1 measure 95.21% with least Mean Squared Error and Mean Absolute Error of 4.02 respectively compared to other algorithms.

5. Conclusion

Cardiovascular disease has emerged in the last ten years as a major medical problem with high death rate. For the intelligent prediction of cardiovascular disease (CVD), combined standard benchmark dataset was utilized to evaluate the efficiency of suggested strategy. Different feature selection methods, such as LASSO, RELIEF, MR-MR, RFE, and PCA were contributed on machine learning algorithms such as Random forest, Logistic Regression, K-Nearest Neighbour, Naïve Bayes, Support Vector Machine, and Neural Network for prediction of heart disease. The outcomes were then compared using assessment criteria such as Accuracy, Precision, Recall, F1score, MSE, and MAE. The primary outcome was thorough comparison of several feature selection strategies on machine algorithms for prediction of cardiac illnesses. The findings revealed that the suggested method computes extremely favourable outcomes for numerous statistically relevant measurement parameters, including F1score, Mean Squared Error, Mean Absolute Error Accuracy, Precision, and Recall. The Random forest and neural network classifier implemented using LASSO and MR-MR feature selection techniques have achieved high Accuracy, Precision, Recall, F1score [43] [44] [45] with least Mean Squared Error and Mean Absolute Error compared with other algorithms. Significant improvements have been made in classification performance. Many real-world applications, such as disease diagnosis, prediction, and engineering optimization issues, may benefit from the use of the suggested approach. This could assist in addressing problems in real life.

References

[1] Abdallah Abdellatif, Hamdan Abdellatef, Jeevan Kanesan, Chee-Onn Chow, Joon Huang Chuah and Hassan Muwafaq Gheni, An Effective Heart Disease Detection and Severity Level Classification Model Using Machine Learning and Hyperparameter Optimization Methods, IEEE Access, VOLUME 10, 2022

[2] G. Bazoukis, S. Stavrakis, J. Zhou, S. C. Bollepalli, G. Tse, Q. Zhang, J. P. Singh, and A. A. Armoundas, "Machine learning versus conventional clinical methods in guiding management of heart failure patients—A systematic review," Heart Failure Rev., vol. 26, no. 1, pp. 23–34, Jan. 2021

[3] Yilmaz R, Yagin FH., "Early detection of coronary heart disease based on machine learning methods", International Medical Journal, 2022 Jan 1; 4(1): 1–6. doi: 10.37990/medr.1011924.

[4] Protima Khan, Md. Fazlul Kader, S. M. Riazul Islam, Aisha B. Rahman, Md. Shahriar Kamal, Masbah Uddin Toha, And Kyung-Sup Kwak, "Machine Learning and Deep Learning Approaches for Brain Disease Diagnosis: Principles and Recent Advances", IEEE Access, VOLUME 9, 2021, Digital Object Identifier 10.1109/ACCESS.2021.3062484

[5] Pronab Ghosh, Sami Azam, Mirjam Jonkman, Asif Karim, F. M. Javed Mehedi Shamrat, Eva Ignatious, Shahana Shultana, Abhijith Reddy Beeravolu, Friso De Boer (2021). Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms with Relief and LASSO Feature Selection Techniques. IEEE Access, Vol9. doi:10.1109/access.2021.3053759

[6] Jian Ping Li, Amin ul Haq, Salah ud Din, Jalaluddin Khan, Asif Khan and Abdus Saboor (2020), Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare. IEEE Access, 8(), 107562–107582. doi:10.1109/ACCESS.2020.3001149

[7] Usama Ahmed, Ghassan F. Issa, Muhammad Adnan Khan, Shabib Aftab, Muhammad Farhan Khan, Raed A. T. Said, Taher M. Ghazal and Munir Ahmad, Prediction of Diabetes empowered with Fused Machine Learning, IEEE Access, VOLUME 10, 2022

[8] Ghulab Nabi Ahmad, Shafiullah, Abdullah Algethami, Hira Fatima, and Syed Md. Humayun Akhter, Comparative Study of Optimum Medical Diagnosis of Human Heart Disease Using Machine Learning Technique with and Without Sequential Feature Selection, IEEE Access, VOLUME 10, 2022

[9] Sarria E. A. Ashri, M. M. El-Gayar , and Eman M. El-Daydamony, Heart Disease Prediction Framework Based on Hybrid Classifiers and Genetic Algorithm, IEEE Access, VOLUME 9, 2021

[10] Huazhong Yang, Zhongju Chen, Huajian Yang, and Maojin Tian, Predicting Coronary Heart Disease Using an Improved LightGBM Model: Performance Analysis and Comparison, IEEE Access, VOLUME 11, 2023

[11] Mohammed B. Abubaker and Bilal Babayigit, Detection of Cardiovascular Diseases in ECG Images Using Machine Learning and Deep Learning Methods, IEEE Transactions on Artificial Intelligence, VOL. 4, NO. 2, APRIL 2023

[12] Tsatsral Amarbayasgalan, Van-Huy Pham, Nipon Theera-Umpon, Yongjun Piao and Keun Ho Ryu, An Efficient Prediction Method for Coronary Heart Disease Risk Based on Two Deep Neural Networks Trained on Well-Ordered Training Datasets, IEEE Access, VOLUME 9, 2021

[13] D. Cenitta, R. Vijaya Arjunan and K. V. Prema, Ischemic Heart Disease Prediction Using Optimized Squirrel Search Feature Selection Algorithm, IEEE Access, VOLUME 10, 2022, Digital Object Identifier 10.1109/ACCESS.2022.3223429

[14] Senthilkumar Mohan, Chandrasegar Thirumalai and Gautam Srivastava, Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques, IEEE Access, VOLUME 7, 2019, Digital Object Identifier 10.1109/ACCESS.2019.2923707

[15] Chunyan Guo, Jiabing Zhang, Yang Liu, Yaying Xie, Zhiqiang Han and Jianshe Yu, Recursion Enhanced Random Forest With an Improved Linear Model (RERF-ILM) for Heart Disease Detection on the Internet of Medical Things Platform, IEEE Access, VOLUME 8, 2020, Digital Object Identifier 10.1109/ACCESS.2020.2981159

[16] Mohammad Ayoub Khan and Fahad Algarnia, Healthcare Monitoring System for the Diagnosis of Heart Disease in the IoMT Cloud Environment Using MSSO-ANFIS, IEEE Access, VOLUME 8, 2020, Digital Object Identifier 10.1109/ACCESS.2020.3006424

[17] Syed Arslan Ali, Basit Raza, Ahmad Kamran Malik, Ahmad Raza Shahid, Muhammad Faheem, Hani Alquhayz and Yogan Jaya Kumar, An Optimally Configured and Improved Deep Belief Network (OCI-DBN) Approach for Heart Disease Prediction Based on Ruzzo–Tompa and Stacked Genetic Algorithm, IEEE Access, VOLUME 8, 2020, Digital Object Identifier 10.1109/ACCESS.2020.2985646

[18] Gihun Joo, Yeongjin Song, Hyeonseung Im, And Junbeom Park, Clinical Implication of Machine Learning in Predicting the Occurrence of Cardiovascular Disease Using Big Data (Nationwide Cohort Data in Korea), IEEE Access, VOLUME 8, 2020, Digital Object Identifier 10.1109/ACCESS.2020.3015757

[19] Ashir Javeed, Shijie Zhou, Liao Yongjian, Iqbal Qasim, Adeeb Noor and Redhwan Nour, An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection, IEEE Access, VOLUME 7, 2019, Digital Object Identifier 10.1109/ACCESS.2019.2952107

[20] Liaqat Ali, Awais Niamat, Javed Ali Khan, Noorbakhsh Amiri Golilarz, Xiong Xingzhong, Adeeb Noor, Redhwan Nour and Syed Ahmad Chan Bukhari, An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure, IEEE Access, VOLUME 7, 2019, Digital Object Identifier 10.1109/ACCESS.2019.2909969

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1445-1471, 2024
DOI: 10.55214/25768484.v8i5.1848

[21] Liaqat Ali, Atiqur Rahman, Aurangzeb Khan, Mingyi Zhou, Ashir Javeed and Javed Ali Khan, An Automated Diagnostic System for Heart Disease Prediction Based on χ 2 Statistical Model and Optimally Configured Deep Neural Network, IEEE Access, VOLUME 7, 2019, Digital Object Identifier 10.1109/ACCESS.2019.2904800
[22] Jikuo Wang, Changchun Liu, Liping Li, Wang Li, Lianke Yao, Han Li and Huan Zhang, A Stacking-Based Model for Non-Invasive Detection of Coronary Heart Disease, IEEE Access, VOLUME 8, 2020, Digital Object Identifier 10.1109/ACCESS.2020.2975377
[23] Abid Ishaq, Saima Sadiq, Muhammad Umer, Saleem Ullah, Seyedali Mirjalili, Vaibhav Rupapara and Michele Nappi, Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques, IEEE Access, VOLUME 9, 2021, Digital Object Identifier 10.1109/ACCESS.2021.3064084
[24] Dakun Lai, Yifei Zhang, Xinshu Zhang, Ye Su and Md Belal Bin Heyat, An Automated Strategy for Early Risk Identification of Sudden Cardiac Death by Using Machine Learning Approach on Measurable Arrhythmic Risk Markers, IEEE Access, VOLUME 7, 2019, Digital Object Identifier 10.1109/ACCESS.2019.2925847
[25] Saba Bashir, Abdulwahab Ali Almazroi, Sufyan Ashfaq, Abdulaleem Ali Almazroi and Farhan Hassan Khan, A Knowledge-Based Clinical Decision Support System Utilizing an Intelligent Ensemble Voting Scheme for Improved Cardiovascular Disease Prediction, IEEE Access, VOLUME 9, 2021, Digital Object Identifier 10.1109/ACCESS.2021.3110604
[26] Aqsa Rahim, Yawar Rasheed, Farooque Azam, Muhammad Waseem Anwar, Muhammad Abdul Rahim and Abdul Wahab Muzaffar, "An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases", IEEE Access, VOLUME 9, 2021, Digital Object Identifier 10.1109/ACCESS.2021.3098688
[27] Shishir Rao, Yikuan Li, Rema Ramakrishnan, Abdelaali Hassaine, Dexter Canoy, John Cleland, Thomas Lukasiewicz, Gholamreza Salimi-Khorshidi, and Kazem Rahimi, "An Explainable Transformer-Based Deep Learning Model for the Prediction of Incident Heart Failure", IEEE Journal of Biomedical and Health Informatics, Vol. 26, No. 7, July 2022
[28] Chayakrit Krittanawong, Hafeez Ul Hassan Virk, Sripal Bangalore, Zhen Wang, Kipp W. Johnson, Rachel Pinotti, HongJu Zhang, Scott Kaplin, Bharat Narasimhan, Takeshi Kitai, Usman Baber, Jonathan L. Halperin & W. H. Wilson Tang," Machine learning prediction in cardiovascular diseases: a meta-analysis", Sci Rep 10, 16057 (2020). https://doi.org/10.1038/s41598-020-72685-1
[29] Malavika G, Rajathi N, Vanitha V and Parameswari P, "Heart Disease Prediction Using Machine Learning Algorithms", Bioscience Biotechnology Research Communication, Special Issue Vol 13 No 11 (2020) Pp-24-27.
[30] Md Mamun Ali, Bikash Kumar Paul, Kawsar Ahmed, Francis M. Bui, Julian M. W. Quinn, Mohammad Ali Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison", Computers in Biology and Medicine, Volume 136, September 2021, 104672.
[31] P. Shashwith, B. Sai Prasad, CH. Shashi Kumar Reddy, N. Abhinav Krishna, K. Ramesh, "Heart Disease Prediction by using Machine Learning Algorithms", International Research Journal of Modernization in Engineering Technology and Science, Volume:04/Issue:06/June-2022.
[32] Sudipta Priyadarshinee and Madhumita Panda, "Cardiac Disease Prediction using Smote and Machine Learning Classifiers", Journal of Pharmaceutical Negative Results ¦ Volume 13 ¦ Special Issue 8 ¦ 2022.
[33] Irfan Javid, Ahmed Khalaf Zager Alsaedi and Rozaida Ghazali, "Enhanced Accuracy of Heart Disease Prediction using Machine Learning and Recurrent Neural Networks Ensemble Majority Voting Method" International Journal of Advanced Computer Science and Applications(IJACSA),Volume11,Issue3,2020. http://dx.doi.org/10.14569/IJACSA.2020.0110369
[34] Md. Julker Nayeem, Sohel Rana, and Md. Rabiul Islam, "Prediction of Heart Disease Using Machine Learning Algorithms", European Journal of Artificial Intelligence and Machine Learning, Vol 1 | Issue 3 | November 2022.
[35] Lijetha.C. Jaffrin, Dr.J. Visumathi, Dr.G.Umarani Srikanth, "Role of Exploratory Analytics and Visualization in Heart Disease Prediction", International Journal of Engineering Trends and Technology Volume 70 Issue 1, 24-34, January, 2022.
[36] Noroozi, Z., Orooji, A. & Erfannia, L., "Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction", Sci Rep 13, 22588 (2023). https://doi.org/10.1038/s41598-023-49962-w.
[37] Kaushalya Dissanayake, Md Gapar Md Johar, "Comparative Study on Heart Disease Prediction Using Feature Selection Techniques on Classification Algorithms", Applied Computational Intelligence and Soft Computing, Volume 2021, Article ID 5581806, https://doi.org/10.1155/2021/5581806.
[38] Bharti R, Khamparia A, Shabaz M, Dhiman G, Pande S, Singh P., "Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning", Computational Intelligence and Neuroscience. 2021 Jul 1; 2021:8387680. doi: 10.1155/2021/8387680. PMID: 34306056; PMCID: PMC8266441.
[39] Xiao-Yan Gao, Abdelmegeid Amin Ali, Hassan Shaban Hassan, Eman M. Anwar, "Improving the Accuracy for Analyzing Heart Diseases Prediction Based on the Ensemble Method", Complexity, 2021 https://doi.org/10.1155/2021/6663455.
[40] Mienye, I.D.; Jere, N., "Optimized Ensemble Learning Approach with Explainable AI for Improved Heart Disease Prediction", Information 2024, 15, 394. https://doi.org/10.3390/info15070394.
[41] Khan Mamun, M.M.R. Elfouly. T., "Detection of Cardiovascular Disease from Clinical Parameters Using a One-Dimensional Convolutional Neural Network", Bioengineering, 2023, 10, 796. https://doi.org/10.3390/bioengineering10070796.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1445-1471, 2024
DOI: 10.55214/25768484.v8i5.1848

[42] Arkadip Ray, Avijit Kumar Chaudhuri, "Smart healthcare disease diagnosis and patient management: Innovation, improvement and skill development", Machine Learning with Applications, Volume 3, 2021, 100011, ISSN 2666-8270, https://doi.org/10.1016/j.mlwa.2020.100011.

[43] Oleiwi, Zahraa Ch., Ebtesam N. AlShemmary, and Salam Al-augby. "Adaptive Features Selection Technique for Efficient Heart Disease Prediction", Journal of Al-Qadisiyah for Computer Science and Mathematics, Vol. 15(1) 2023, ppComp.1–18 https:// doi.org/10.29304/JQCM.2023.15.1.1137

[44] Veena S T, Jeevetha R, Abirami N, "Prediction of Heart Disease using Hybrid Feature Selection", Journal of Positive School Psychology, 2022, Vol. 6, No. 4, 10454–10469.

[45] N. Ghaniaviyanto Ramadhan, Adiwijaya, W. Maharani and A. Akbar Gozali, "Chronic Diseases Prediction Using Machine Learning With Data Preprocessing Handling: A Critical Review," IEEE Access, vol. 12, pp. 80698-80730, 2024, doi: 10.1109/ACCESS.2024.3406748.

[46] Kim, K. ., & Park, H., "Neural network approach to predict the association between blood cadmium levels and hypertension", Edelweiss Applied Science and Technology, 2024, 8(4), 336–344. https://doi.org/10.55214/25768484.v8i4.978

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1445-1471, 2024
DOI: 10.55214/25768484.v8i5.1848