Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 03, No. 02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.32.14.32

Performance Evaluation of Machine Learning Algorithm in Various Datasets

Md. Siraj-Ud-Doulah[1*], Md. Nazmul Islam[2]

[1]Associate Professor, Department of Statistics, Begum Rokeya University, Rangpur, Bangladesh*
[2]M.Sc, Department of Statistics, Begum Rokeya University, Rangpur, Bangladesh

*Corresponding Email: [1]*sdoulah_brur@yahoo.com*

Received: 30 October 2022 **Accepted:** 17 January 2023 **Published:** 21 February 2023

Abstract: Machine learning is one of the fast-growing areas of computer science, with far-reaching applications. There are several applications for machine learning. The most significant of which is supervised learning. Supervised learning is common in classification problems. In this study, frequently used twelve machine learning algorithms are considered: NB, LDA, LR, ANN, SVM, K-NN, HT, DT, C4.5, CART, RF and BB. We apply these algorithms on seven datasets. The main goal of this study was to evaluate the performance of the machine learning algorithms on both binary and multiple classification problems using a variety of performance metrics: accuracy, kappa statistic, precision, recall, specificity, F-measure, MAE, RMSE and MCC. Here, we found that RF algorithm proved to have the best performance in three out of seven datasets. But the other four algorithms: NN, NB, BB and LR also performed well.

Keywords: Machine Learning, Classification, Confusion Matrix, Performance Measures.

1. INTRODUCTION

Machine learning [2,3] can be used in a variety of ways for algorithm selection. A few papers compare different machine learning algorithms in detail. J48, CART, and ADTree are compared by [10]. CART was found to be the most accurate algorithm, with a 98.5 percent accuracy rating. GRU-SVM, LR, MLP, NN, SR, and SVM for breast cancer dataset were used to discover Agarap [1]. MLP was determined to have the highest classification accuracy of 99.04 percent. J48, Nave Bayes, LMT, REP Tree, DT, K-star, LR, ICO, IBK, and FC are used to analyze several machine learning methods for the breast cancer dataset in [4,10,11,12, 13, 14]. Clean classifiers are good with a classification accuracy of 76 percent. For a breast cancer dataset, [14] examine LR, NB, and SVM. They discovered that SVM is the most effective. Sadhana et al. [16] compared DT and SVM on a breast cancer dataset and discovered that the SVM had the highest accuracy of 96.99 percent. Doulah [17] compare three

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 03, No. 02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.32.14.32

algorithms SVM, K-NN, DT and found that SVM grants the largest accuracy of (98.1%). Dana et al. [15] examine five algorithms SVM; ANN; DT; NB; and K-NN. They located that SVM reap the best accuracy is ninety%. Helwan et al. [3] discover the use of lower BPNN and RBFN for breast tissue dataset and locate RBFN outperforms the BPNN in phrases of accuracy. Ravi [18] assessed NB, IBK and J48 to procedure classify the Breast Tissue dataset. They discovered that the J48 set of rules proficiency is higher than other algorithms. Subramani et al. [20] compare 3 classification algorithms are SVM, KNN and DT and discovered that SVM performs. Kleyko et al. [21] examined algorithms NN, LR, SVM and determined that LR is the pleasant choice with 93.4%. Thomas [19] compares BB, DT, C4.5 and found that BB is outperform. Ansari et al. [22] examined NN, BB, NB and found that NB, the highest accuracy is 93.87%. [24] compares NB, NN, SMO, IBK, J48 and RIPPER, BB set of rules on various datasets. [23] compare BPNN, RST-GA, RST-JA and discovered that BPNN performs higher than the opposite algorithms. [21] evaluate the Quest, univariate splits, C4.5, Ind-Cart, and LR and discovered that C4.5 provides quality results.

2. METHODS AND MATERIAL

Naive Bayes (NB)

Naive Bayes classifier [25] considers that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors is given by-

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (1)$$

Where, $P(x|c) = P(x_1|c) \times P(x_2|c) \times \cdots\cdots \times P(x_n|c) \times P(c)$

Where, Equation (1) is the posterior probability; $P(c)$, is the prior probability of class; $P(x|c)$, is the likelihood.

Logistic Regression (LR)

Let's assume a target variable Y and $X = x_1 + x_2 + \ldots\ldots + x_j$ where j is the number of independent variables, then the conditional mean of Y is given x is given by:

$$\pi(x) = E(\frac{Y}{x}) \quad (2)$$

(2) is the expected value of the target variable for a given X. The LR [26] is given by the formula:

$$\pi(x) = \frac{e^{g(x)}}{1+e^{g(x)}} \quad (3)$$

Where, $g(x) = ln\frac{\pi(x)}{1-\pi(x)} = \beta_0 + \beta_1 x_1 + \ldots\ldots + \beta_k x_k$

Equation (3) is called the logit transformation, where$\beta_0, \beta_1, \ldots\ldots, \beta_k$ are estimated using maximum likelihood. The log likelihood is given by: $L(\frac{\beta}{x)=ln\left[l(\frac{\beta}{x})\right]} = \sum_{i=1}^{n}\{y_i\, ln[\pi(x_i)] + (1-y_i)\,\{y_i\, ln[1-\pi(x_i)]\}$.

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 03, No. 02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.32.14.32

Linear Discriminant Analysis (LDA)

LDA was developed in 1936 by R.A. Fisher [27]. We assume the density function$P(X = x|Y = k)$ of *X* for an observation that comes from the *k*th class is known, then by Bayes Theorem we have.

$$P\left(\frac{Y=k}{X=x}\right) = \frac{P\left(\frac{X=x}{Y=k}\right)P(Y=k)}{\sum_{j=1}^{N} P\left(\frac{X=x}{Y=j}\right)P(Y=j)}$$

In general, $f_k(X) = P(X = x|Y = k)$ is the densities. For example, if we assume $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)$ is drawn from a multivariate Gaussian distribution$N(\mu_k, \Sigma)$, then

$$f_k(x) = \frac{1}{(2\pi)^{\frac{p}{2}}|\sum|^{\frac{1}{2}}} exp\left\{-\frac{1}{2}(x-\mu_k)^T\sum^{-1}(x-\mu_k)\right\}$$

The LDA classifier assigns an observation $X = x$ to the class for which

$$\delta_k(x) = x^T\sum^{-1}\mu_k - \frac{1}{2}\mu_k^T\sum^{-1}\mu_k + ln\,\pi_k$$

Is the largest. From here we can see that the decision boundary for LDA is linear.

Support Vector Machine (SVM)

SVM [28, 29] have been added by using Vapnik and collaborators in 1992. According to [30], the selection floor by using SVM for linearly separated space is a hyper-plane and is given below:

$$w \bullet x + b = 0$$

Where x is an arbitrary feature vector,$b$and$w$are learned from training set linearly separable data. The dot product $w \bullet x$is defined by,

$$w \bullet x = \sum_{i=1}^{n} w_i x_i$$

This linear classifier is represented by the hyper-plane $H(w \bullet x + b = 0)$ and defines a region for class +1 patterns$(w \bullet x + b > 0)$ and another region for class -1 pattern $(w \bullet x + b < 0)$.

Classification and Regression Tree (Cart)

CART algorithm changed into advanced through Brieman, Friedman, Olshen, and Stone in 1984 [32]. For a binary magnificence the GINI degree of impurity is given by, $GINI(t) = 1 - \sum[p(\frac{t}{j})]^2$

Where$p(\frac{t}{j})$ is the relative frequency of class j at node t. When a node p is split into x partitions, the quality of split is given by

$$GINI_{split} = \sum_{i=1}^{n} \frac{n_i}{n} GINI$$

Wheren_i=number of records at child$i$

n = number of records at node p

C4.5

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 03, No. 02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.32.14.32

The first step in C4.5 set of rules [33] is to specify root of the tree. To determine the order of functions inside the choice tree, records benefit components is evaluated for each attribute as described in

$$Entropy(S) = \sum_{i=1}^{n} -p_i$$

Where i is a state, p_i is the possibility of outcome being in state i for the set S and n is the number of possible.

K-Nearest Neighbor (K-NN)

KNN [34] is distance-primarily based classifier wherein distance is used to classify information based on labels of its associates which can be decided on from training data.

$$\text{Euclidian Distance: } \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

If k = 1, the class where its nearest neighbor belongs. Alternatively, if we provide a huge ok input, it can bring about underneath-fitting.

Random Forest (RF)

Random Forests [34] similarly weighted is offered as inside the following system:

$$H(x) = arg\left\{max\frac{1}{m}\sum_{i=1}^{m}\left(I(x;\Theta_i) = y_j\right)\right\}$$

$$mg(X,Y) = av_n I(I_k(X) = Y) - \max_{j \neq Y_k}(I_k(X) = j)$$

The generalization error is given by,

$$PE^* = P_{X,Y}(mg(X,Y) < 0)$$

The strength of the set of classifiers$\{I(x,\Theta)\}$is

$$s = E_{X,Y}\left(P_\Theta(I(x,\Theta) = Y) - \max_{j \neq Y_\Theta}(I(x,\Theta) = j)(2)\right)$$

An upper bound for the generalization error is given by

$$PE^* \leq \frac{\overline{\rho}(1-s^2)}{s^2}$$

Bagging and Boosting (BB)

Let $D = \{(x_1,y_1),(x_2,y_2),\ldots\ldots,(x_m,y_m)\}$be a set of $m$items and let$y_i \in Y = \{c_1,c_2,\ldots\ldots,c_k\}$be a set of *k* class labels; *C* denotes classification algorithm and *n* is number of learners.

1. Draw *m* items randomly with replacement from the dataset *D*, so generate bootstrap samples, $D_1, D_2,\ldots,D_n$
2. Each dataset D_i is trained and multiple classification models are constructed, $M_i = C(D_i)$.
3. Consensus of classification models is tested to calculate out-of-bag error.
4. New sample *x* is given to classifiers as input and the outputs y_i are obtained from each model $y_i = M_i(x)$
5. The outputs of models $\{M_1, M_2,\ldots,M_t\}$ are combined as in

$$M^*(x) = \underset{y \in Y}{argmax}\sum_{i:M_i(x)=y}^{n} 1$$

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 03, No. 02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.32.14.32

Hoeffding Tree (HT)

The Hoeffding Bound [31,35,36] is calculated the usage of the following equation: Hoeffding bound**,**

$€=\sqrt{\frac{R^2\, ln(\frac{1}{\delta})}{2nl}}$

As mentioned before, the Hoeffding Tree algorithm incrementally generates a decision tree from Data Streams.

Decision Table (DT)

Decision table is an accurate technique for making numerical predictions from decision trees. It is an ordered set of If-Then rules, which may be denser than decision trees and therefore easier to understand [37].

Evaluation Criteria

Table 1 shows a confusion matrix illustrating what might happen for positive and negative outcomes in a data mining model.

Table 1. Confusion Matrix

	Predicted	
Actual	TP	FN
	FP	TN

The performance criterions are shown in the following Table 2:

Table 2. Performance Criterions

Performance Criterions	**Statistics**
Accuracy	$A=\frac{TP+TN}{TP+TN+FP+FN}$
Cohen's kappa	$\kappa = \frac{p_o-p_e}{1-p_e} = 1-\frac{1-p_o}{1-p_e}$, Where, $P_o = \frac{TP+TN}{TP+TN+FP+FN}$ and $P_e = (\frac{TP+FN}{TP+TN+FP+FN}.\frac{TP+FP}{TP+TN+FP+FN}) + (\frac{FP+TN}{TP+TN+FP+FN}.\frac{FN+TN}{TP+TN+FP+FN})$
Precision	$P = \frac{TP}{TP+FP} = \frac{C_{ii}}{\sum_{k=1}^{n+1} C_{ki}} = \frac{C_{ii}}{PTotal_i}$, Where $PTotal_i$ is the Predicted total for the class and C_{ii} is the Predicted class for a multi class problem.
Recall or Sensitivity	$R/S = \frac{TP}{FN+TP} = \frac{C_{ii}}{\sum_{i=1}^{n+1} C_{ij}} = \frac{C_{ii}}{ATotal_i}$, Where $ATotal_i$ is the total for the actual total and C_{ii} is the Predicted class for a multi class problem, and n is the sample size.
Specificity	$S = \frac{TN}{TN+FP} = \frac{T-ATotal_i-PTotal_i+C_{ii}}{T-ATotal_i}$
F-Measure	$F = \frac{2*Recall*Precision}{Recall+Precision}$

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 03, No. 02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.32.14.32

Matthews's Correlation Coefficient	$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)\,(TP+FN)\,(TN+FP)\,(TN+FN)}}$
Mean Absolute Error	$MAE = \frac{1}{n}\sum_{j=1}^{n}\lvert y_j - \hat{y}_j \rvert$
Root Mean Square Error	$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$

Software Used

The open-source programming language R is used to analyze statistical data, present graphics, and generate reports [38]. The experiments were conducted using open-source R software, version 3.5.2 (https://www.r-project.org).

Dataset Description

The summary of the datasets used in comparative studies in the following Table 3.

Table 3. Dataset Characteristics

Sl. No	Dataset	No. of Attributes (Including Class)	No. of Instances	No of Classes	Missing Values
1	Breast Cancer Wisconsin (Original)	11	699	2	Yes
2	Statlog (Vehicle Silhouettes)	19	846	4	No
3	Vertebral Column	7	310	3	No
4	Breast Tissue	10	106	6	No
5	Contraceptive method Choice	10	1473	3	No
6	Image segmentation	20	2310	7	No
7	Artificial Characters	8	10218	10	No

The datasets chosen for the article have been downloaded from UCI repository, https://archive.ics.uci.edu/ml/datasets/ and https://www.openml.org/search?type=data.

Missing Value Estimation

There are numerous methods available for estimating missing values [40], to estimate missing values for our analysis, however, we use mode method.

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 03, No. 02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.32.14.32

Hold Out Approach

The hold out approach is a straightforward method for generating test data. This method divides the data set into two subsets, a training set and a test set, at random. In this study, the training set will contain 80% of the data, while the test set will contain the remaining 20%. We use the training data set to train the model and use the test data set to test the model based on the training model.

3. RESULTS AND DISCUSSIONS

A simple and sensible way to start is by looking at the data frame, using pairs to plot every variable against every other. There appears to be excellent data separation, and reasonable separation, but nothing obvious for the other variables. The commonest plots for a sample are histograms and box plots. Histograms are excellent for showing the mode, the spread, and the symmetry (skew) of a set of data [8] whereas boxplots are outstanding for showing the spread [9], outliers [6,7] and normality of a set of data [5]. From Figure 1, we found that the Wisconsin breast cancer dataset contains two classes. From Figure 2, we visualize that the breast tissue dataset contains six classes. From Figure 3, we envisage that the vehicle dataset contains four classes. From Figure 4, we found that the vertebral column dataset contains three classes. From Figure 5, we visualize that the contraceptive method choice dataset contains three classes. From Figure 6, we visualize that the image segmentation dataset contains seven classes. From Figure 7, we envisage that the artificial characters dataset contain ten classes and we also see the correlation coefficient between the attribute

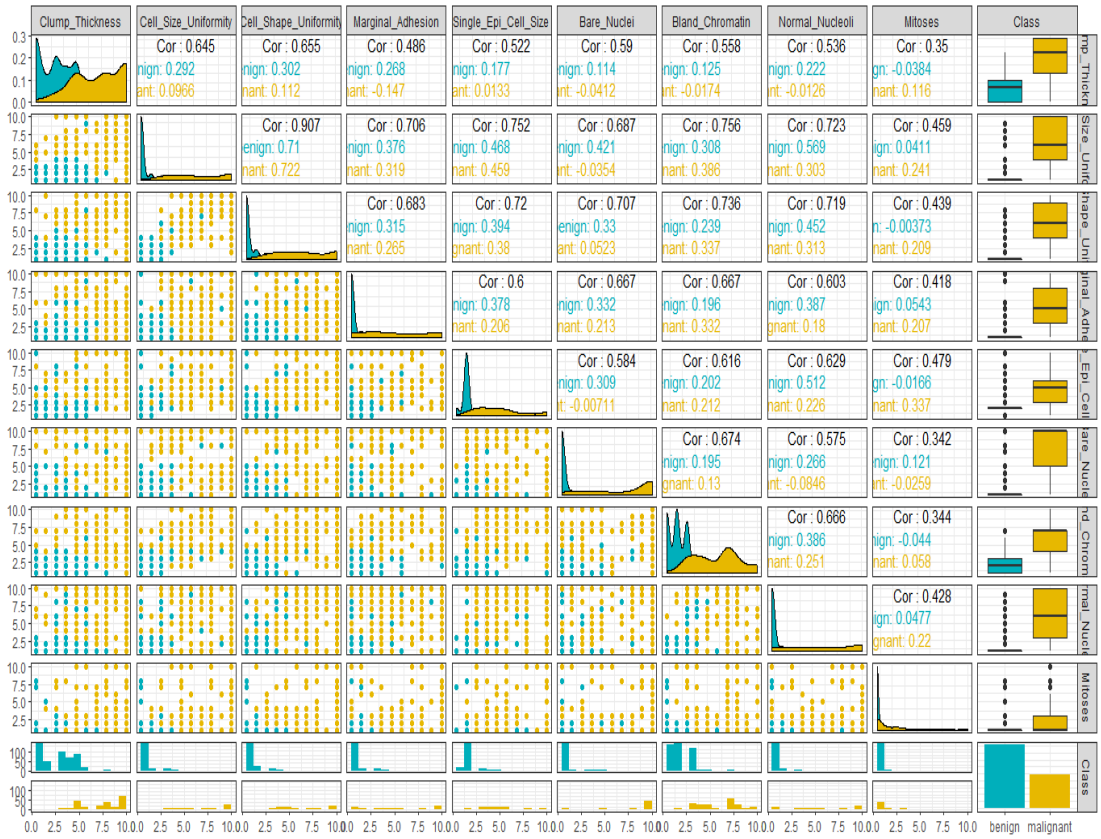

Figure 1. Scatterplot Matrix by Groups for Breast Cancer Dataset

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 03, No. 02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.32.1.13

Table 4. Performance Measures of Breast Cancer Dataset

CA	CCI (Accuracy)	ICI	Kappa	MAE	RMSE	Precision	Recall	Specificity	F-measure	MCC
NB	96.43%	3.57 %	91.77 %	0.035	0.187	0.966	0.964	0.971	0.965	0.919
LDA	95%	5%	88.02 %	0.047	0.196	0.950	0.950	0.913	0.949	0.881
LR	97.14%	2.85 %	93.37 %	0.041	0.163	0.972	0.971	0.974	0.972	0.971
NN	96.42%	3.57 %	91.77 %	0.039	0.155	0.966	0.964	0.971	0.965	0.919
SVM	97.14%	2.85 %	93.37 %	0.063	0.169	0.972	0.971	0.974	0.972	0.934
KNN	95%	5%	88.33 %	0.047	0.211	0.950	0.950	0.939	0.950	0.883
BG	95.71%	4.29 %	89.93 %	0.072	0.178	0.957	0.957	0.942	0.957	0.899
DT	95%	5%	88.33 %	0.078	0.193	0.950	0.950	0.939	0.950	0.883
C4.5	93.57%	6.43 %	84.99 %	0.079	0.243	0.936	0.936	0.920	0.936	0.850
RF	97.85%	2.14 %	95.03 %	0.032	0.150	0.980	0.979	0.991	0.979	0.952
CART	93.57%	6.43 %	84.99 %	0.101	0.239	0.936	0.936	0.920	0.936	0.850
HT	96.43%	3.57 %	91.77 %	0.036	0.188	0.966	0.964	0.971	0.965	0.919

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 03, No. 02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.32.14.32

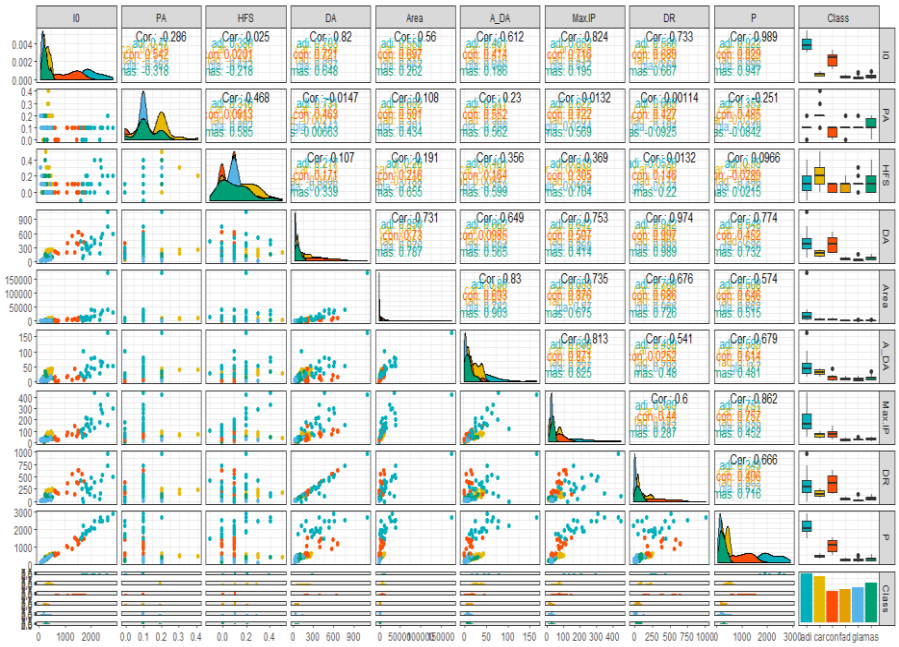

Figure 2. Scatterplot Matrix by Groups for Breast Tissue Dataset

Table 5. Performance Measures of Breast Tissue Dataset

CA	CCI (Accuracy)	ICI	Kappa	MAE	RMSE	Precision	Recall	Specificity	F-measure	MCC
NB	54.55%	45.45 %	45%	0.124	0.315	0.633	0.545	0.926	0.579	0.497
LDA	59.09%	40.91 %	51.11 %	0.157	0.309	0.735	0.591	0.948	0.629	0.577
LR	68.18%	31.81 %	61.79 %	0.112	0.294	0.759	0.682	0.958	0.692	0.650
NN	63.63%	36.36 %	57.18 %	0.137	0.279	0.740	0.636	0.971	0.684	0.621
SVM	50%	50%	42.92 %	0.243	0.342	0.640	0.500	0.976	0.561	0.490
KNN	77.27%	22.73 %	72.29 %	0.089	0.267	0.909	0.773	0.978	0.824	0.800
BG	63.63%	36.36 %	56%	0.145	0.289	0.682	0.636	0.925	0.657	0.577
DT	59.09%	40.91 %	49.87 %	0.189	0.289	0.645	0.591	0.934	0.604	0.540
C4.5	72.73%	27.27 %	66.67 %	0.093	0.289	0.806	0.727	0.960	0.756	0.712
RF	63.64%	36.36 %	56.86 %	0.130	0.268	0.703	0.636	0.964	0.646	0.605

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 03, No. 02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.32.14.32

CART	63.64%	36.36 %	56%	0.147	0.318	0.682	0.636	0.925	0.657	0.577
HT	59.09%	40.91 %	49.87 %	0.128	0.323	0.648	0.591	0.923	0.617	0.531

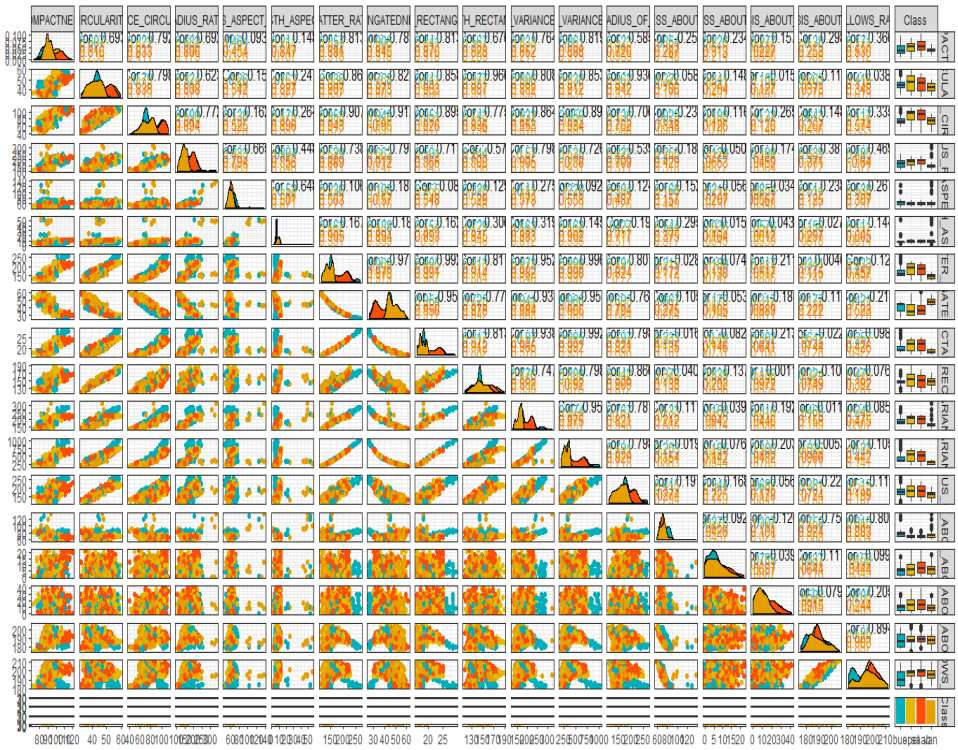

Figure 3. Scatterplot Matrix by Groups for Vehicle Dataset

Table 6. Performance Measures of Vehicle Dataset

CA	CCI	ICI	Kappa	MAE	RMSE	Precision	Recall	Specificity	F-measure	MCC
NB	44.70%	55.25%	27.41%	0.276	0.451	0.520	0.447	0.828	0.401	0.290
LDA	77.05%	22.94%	69.39%	0.137	0.287	0.758	0.771	0.923	0.758	0.686
LR	81.17%	18.82%	74.87%	0.116	0.272	0.804	0.812	0.934	0.805	0.744
NN	78.82%	21.17%	71.73%	0.124	0.302	0.778	0.788	0.928	0.782	0.711
SVM	72.35%	27.64%	63.18%	0.282	0.359	0705	0.724	0.907	0.709	0.621
KNN	68.23%	31.76%	57.61%	0.160	0.397	0.665	0.682	0.891	0.672	0.565
BG	71.17%	28.83%	61.56%	0.167	0.288	0.687	0.712	0.902	0.695	0.602
DT	63.52%	36.47%	51.57%	0.245	0.343	0.610	0.635	0.880	0.609	0.501
C4.5	68.82%	31.18%	58.44%	0.162	0.363	0.692	0.688	0.895	0.688	0.583
RF	71.76%	28.23%	62.4%	0.158	0.282	0.705	0.718	0.905	0.705	0.615
CART	74.71%	25.29%	66.28%	0.157	0.303	0.733	0.747	0.914	0.738	0.654
HT	45.29%	54.70%	28.16%	0.274	0.448	0.530	0.453	0.830	0.411	0.300

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 03, No. 02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.32.14.32

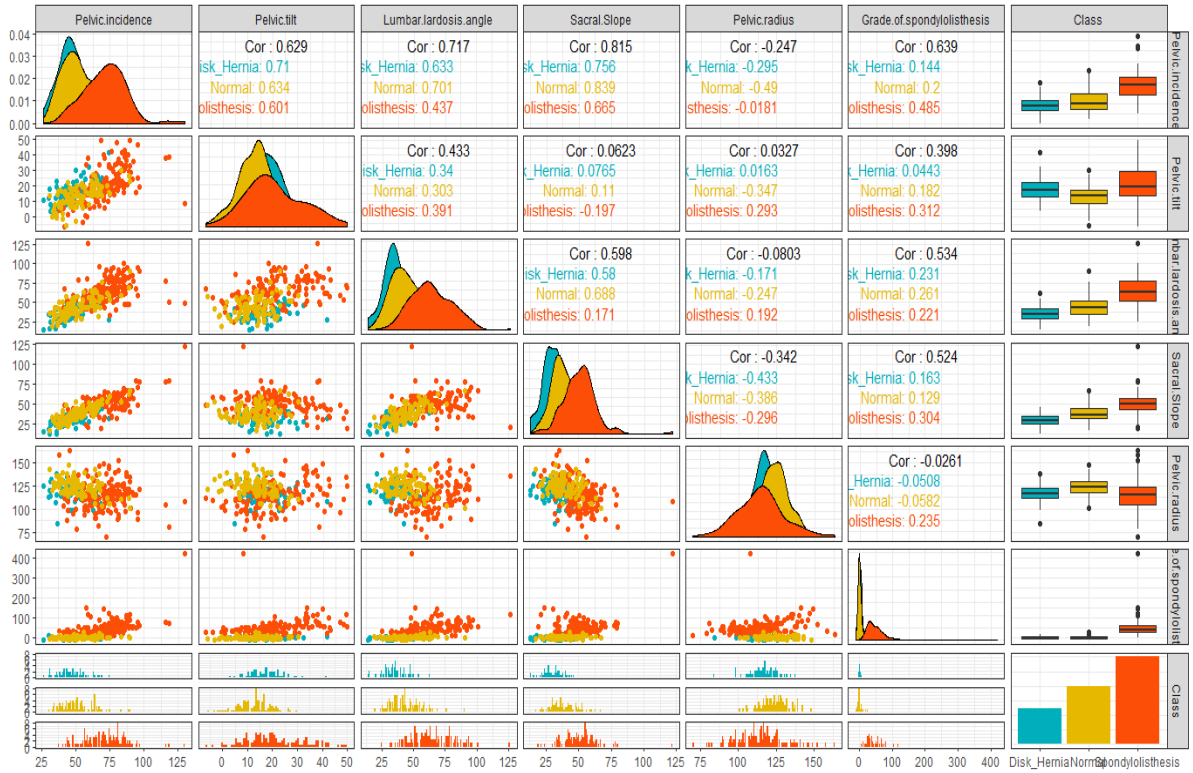

Figure 4. Scatterplot Matrix by Groups for Vertebral Column Dataset

Table 7. Performance Measures of Vertebral Column Dataset

CA	CCI	ICI	Kappa	MAE	RMSE	Precision	Recall	Specificity	F-measure	MCC
NB	83.87%	16.13%	73.24%	0.123	0.262	0.834	0.839	0.934	0.835	0.767
LDA	80.64%	19.35%	68.82%	0.195	0.304	0.830	0.806	0.913	0.813	0.715
LR	88.71%	11.29%	81.38%	0.102	0.224	0.896	0.887	0.951	0.886	0.843
NN	88.71%	11.29%	81.44%	0.097	0.215	0.899	0.887	0.953	0.887	0.844
SVM	70.96%	29.03%	51.98%	0.294	0.380	0.753	0.710	0.855	0.688	0.579
KNN	69.35%	30.64%	50.92%	0.207	0.449	0.734	0.694	0.864	0.694	0.558
BG	87.10%	12.90%	78.71%	0.128	0.245	0.885	0.871	0.946	0.870	0.824
DT	77.42%	22.58%	62.34%	0.196	0.312	0.853	0.774	0.902	0.745	0.704
C4.5	87.10%	12.90%	78.74%	0.110	0.265	0.871	0.871	0.947	0.871	0.818
RF	87.10%	12.90%	78.71%	0.133	0.258	0.885	0.871	0.946	0.870	0.824
CART	87.10%	12.90%	78.71%	0.124	0.266	0.885	0.871	0.946	0.870	0.824
HT	83.87%	16.13%	73.24%	0.121	0.258	0.834	0.839	0.924	0.835	0.767

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 03, No. 02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.32.14.32

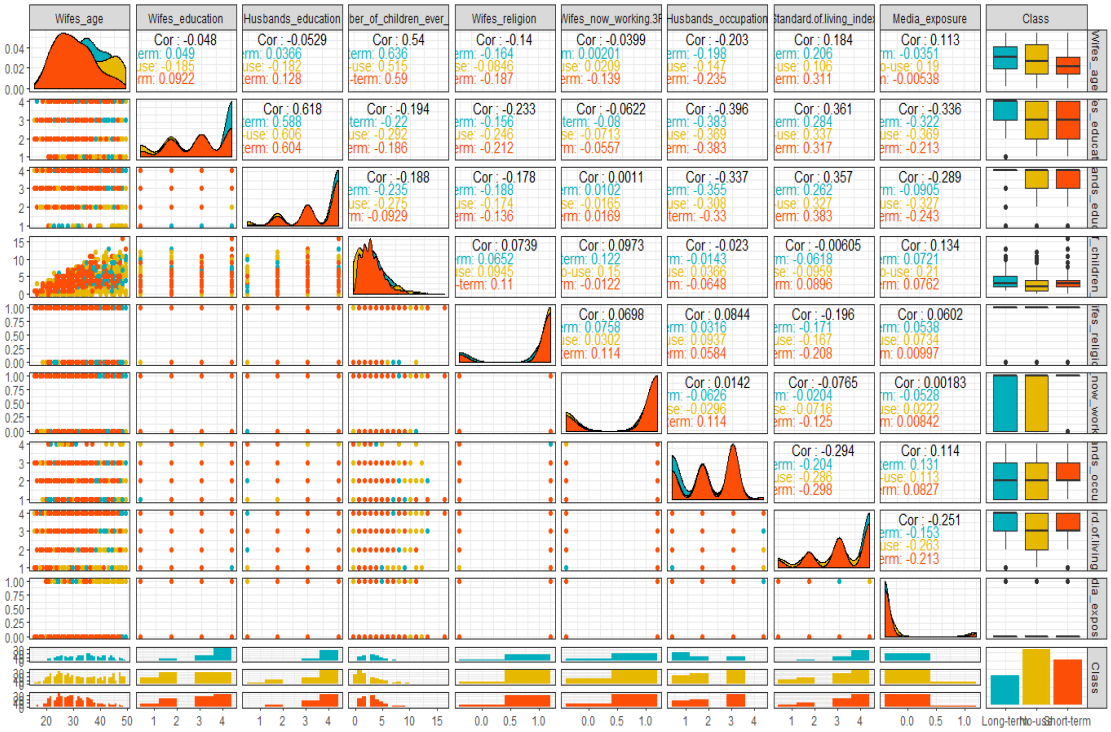

Figure 5. Scatterplot Matrix by Groups for CMC Dataset

Table 8. Performance Measures of CMC Dataset

CA	CCI	ICI	Kappa	MAE	RMSE	Precision	Recall	Specificity	F-measure	MCC
NB	44.74%	55.25%	17.06%	0.382	0.476	0.473	0.447	0.731	0.453	0.179
LDA	53.56%	46.44%	26.92%	0.385	0.443	0.526	0.536	0.743	0.525	0.282
LR	51.86%	48.13%	23.86%	0.382	0.442	0.501	0.519	0.729	0.503	0.250
NN	52.88%	47.12%	28.27%	0.360	0.447	0.549	0.545	0.758	0.529	0.305
SVM	50.51%	49.49%	22.83%	0.372	0.473	0.502	0.505	0.732	0.500	0.239
KNN	41.02%	58.98%	9.24%	0.391	0.617	0.416	0.410	0.684	0.411	0.096
BG	54.24%	45.76%	28.9%	0.330	0.431	0.566	0.592	0.765	0.536	0.305
DT	53.90%	46.10%	27.27%	0.381	0.439	0.541	0.539	0.747	0.532	0.292
C4.5	53.22%	46.78%	26.77%	0.349	0.469	0.525	0.532	0.743	0.522	0.277
RF	50.51%	49.49%	22.19%	0.355	0.451	0.497	0.505	0.725	0.496	0.234
CART	51.53%	48.47%	25.86%	0.352	0.435	0.543	0.515	0.756	0.517	0.279
HT	43.39%	56.61%	16.63%	0.391	0.506	0.470	0.434	0.737	0.434	0.170

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 03, No. 02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.32.14.32

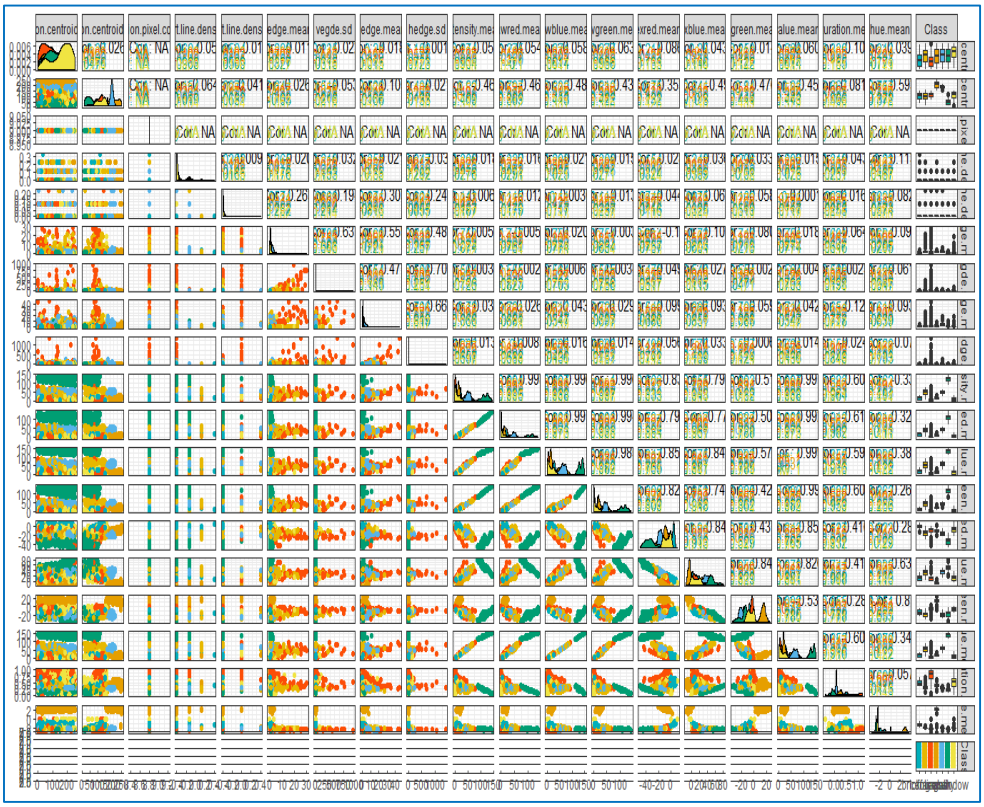

Figure 6. Scatterplot Matrix by Groups for Image Segmentation Dataset

Table 9. Performance Measures of Image Segmentation Dataset

CA	CCI	ICI	Kappa	MAE	RMSE	Precision	Recall	Specificity	F-measure	MCC
NB	79.43%	20.56%	75.93%	0.061	0.234	0.810	0.794	0.964	0.776	0.757
LDA	91.12%	8.87%	89.64%	0.03	0.142	0.912	0.911	0.985	0.910	0.896
LR	96.10%	3.89%	95.45%	0.017	0.094	0.963	0.961	0.993	0.961	0.955
NN	97.62%	2.38%	97.22%	0.04	0.082	0.977	0.976	0.996	0.976	0.972
SVM	93.07%	6.92%	91.91%	0.205	0.303	0.931	0.931	0.988	0.930	0.918
KNN	97.40%	2.60%	96.97%	0.017	0.086	0.974	0.974	0.995	0.974	0.969
BG	96.97%	3.03%	96.46%	0.020	0.085	0.970	0.970	0.995	0.970	0.965
DT	89.83%	10.17%	88.19%	0.097	0.178	0.930	0.898	0.984	0.904	0.895
C4.5	96.32%	3.67%	95.7%	0.122	0.099	0.963	0.963	0.994	0.963	0.957
RF	98.48%	1.51%	98.23%	0.014	0.069	0.985	0.985	0.997	0.985	0.982
CART	96.53%	3.46%	95.95%	0.017	0.091	0.966	0.965	0.994	0.965	0.960
HT	79%	21%	75.44%	0.064	0.238	0.795	0.790	0.963	0.774	0.750

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 03, No. 02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.32.14.32

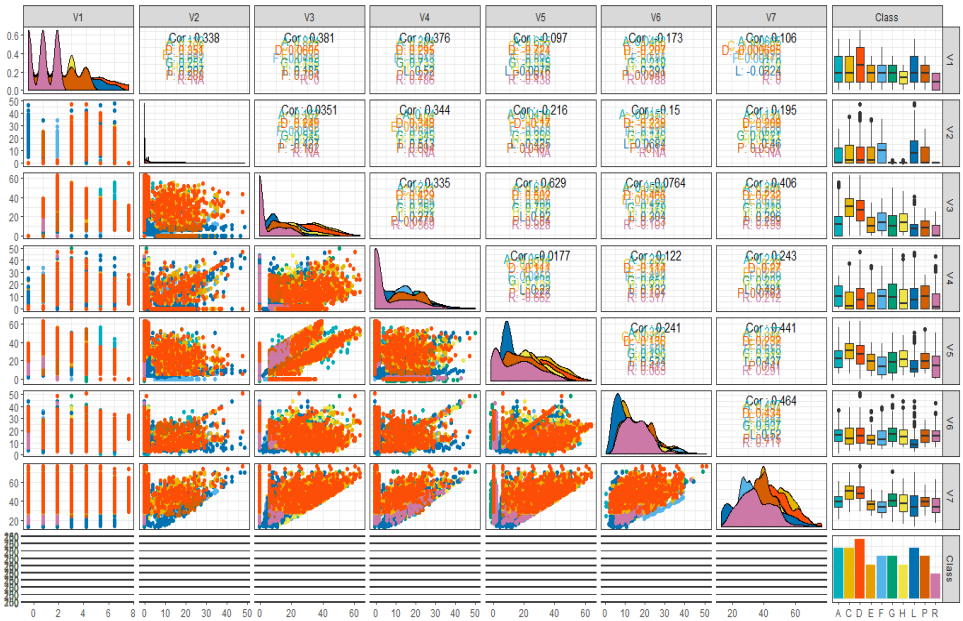

Figure 7. Scatterplot Matrix by Groups for Artificial Characters Dataset

Table 10. Performance Measures of Artificial Characters Dataset

CA	CCI	ICI	Kappa	MAE	RMSE	Precision	Recall	Specificity	F-measure	MCC
NB	29.30%	70.69%	21.88%	0.15	0.296	0.356	0.293	0.928	0.289	0.235
LDA	34.73%	65.26%	27.04%	0.152	0.277	0.351	0.347	0.923	0.334	0.269
LR	34.83%	65.16%	27.19%	0.149	0.273	0.334	0.348	0.924	0.333	0.263
NN	50.10%	49.90%	44.33%	0.113	0.244	0.610	0.501	0.942	0.481	0.469
SVM	40.95%	59.05%	34.11%	0.167	0.284	0.424	0.409	0.932	0.401	0.342
KNN	90.31%	9.68%	89.18%	0.019	0.136	0.904	0.903	0.989	0.903	0.892
BG	74.41%	25.59%	71.44%	0.071	0.184	0.748	0.744	0.971	0.744	0.717
DT	65.02%	34.98%	60.81%	0.120	0.223	0.669	0.650	0.958	0.642	0.612
C4.5	73.73%	26.27%	70.7%	0.053	0.192	0.744	0.737	0.971	0.737	0.709
RF	**91.10%**	**8.90%**	**90.06%**	**0.019**	**0.132**	**0.912**	**0.911**	**0.990**	**0.911**	**0.901**
CART	68.44%	31.56%	64.81%	0.070	0.206	0.691	0.684	0.965	0.685	0.651
HT	35.81%	64.18%	28.91%	0.135	0.306	0.481	0.358	0.933	0.340	0.316

Table 11. The Overall Performance

Dataset	CA	CCI	ICI	Kappa	MAE	RMSE	Precision	Recall	Specificity	F-measure	MCC
Breast cancer	**RF**	97.85%	2.14%	95.03%	0.032	0.150	0.980	0.979	0.991	0.979	0.952

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 03, No. 02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.32.14.32

Breast Tissue	**KNN**	77.27%	22.73%	72.29%	0.089	0.267	0.909	0.773	0.978	0.824	0.800
Vehicle	**LR**	81.17%	18.82%	74.87%	0.116	0.272	0.804	0.812	0.934	0.805	0.744
Vertebral column	**NN**	88.71%	11.29%	81.44%	0.097	0.215	0.899	0.887	0.953	0.887	0.844
CMC	**BG**	54.24%	45.76%	28.9%	0.330	0.431	0.566	0.592	0.765	0.536	0.305
Image segmentation	**RF**	98.48%	1.51%	98.23%	0.014	0.069	0.985	0.985	0.997	0.985	0.982
Artificial characters	**RF**	91.10%	8.90%	90.06%	0.019	0.132	0.912	0.911	0.990	0.911	0.901

For Table 4, all algorithms provide relatively higher accuracy and random forest gives the highest accuracy (97.85%). It is worth noting that all algorithms provide relatively higher kappa value and random forest gives the highest value (95.03%). It is worth mentioned that all algorithms provide relatively higher precision and random forest gives the highest value (98%). It is notice that all algorithms provide relatively higher recall and random forest gives the highest value (97.9%). From the table, all algorithms provide relatively higher specificity and random forest gives the highest value (99.1%).

For Table 5, all the algorithm provides average performance and KNN has the highest accuracy (77.27%). all of the algorithm provides average performance and C4.5 has the highest kappa statistic (66.67%). It is worth noting that all the algorithm provides good performance and KNN has the highest precision value of (90.9%). It is noticed that all the algorithms provide good performance and KNN has the highest recall value of (77.3%). It is worth noting that all the algorithm provides good performance and KNN has the highest specificity value of (97.8%).

For Table 6, logistic regression provides the highest accuracy (81.17%) and other algorithms give relatively lower accuracy than logistic regression. It is worth noting that logistic regression provides the highest kappa value (74.87%) and other algorithms give relatively lower value than logistic regression. It is notice that logistic regression provides the highest precision value (80.40%) and other algorithms gives relatively lower value than logistic regression. it can be seen that logistic regression provides the highest recall value (81.2%) and other algorithms give relatively lower value than logistic regression. It is noticed that logistic regression provides the highest specificity value (93.4%) and other algorithms give relatively lower value than logistic regression.

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 03, No. 02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.32.14.32

For Table 7, NN and LR provide highest accuracy (88.71%) and Bagging, C4.5, RF, CART provide the same value (87.10%). It is notice that NN and LR provides highest accuracy (81.38%) and Bagging, C4.5, RF and CART has the kappa value (78.71%). In addition, NN and LR provide the highest precision value (89.60%) and Bagging, RF and CART have the precision of (88.50%). Moreover, NN and LR provide the highest recall value (88.70%) and Bagging, C4.5, RF and CART have the recall value of (87.1%). It is clear that NN provides the highest specificity value (95.3%).

For Table 8, the accuracy of all algorithms is less than 60%. It is noticed that the kappa value of all algorithms is less than 30%. It is worth noting that the precision of all algorithms is less than 60% and NN has the highest precision (56.6%). It is more notice that the recall of all algorithms is below 60% and NN has the highest precision (59.2%). It is worth mentioned that the specificity of NN (76.5%) algorithm are higher than the other algorithms.

In Table 9, all algorithms provide relatively higher accuracy and random forest gives the highest accuracy (98.48%). It is worth noting that all algorithms provide relatively higher kappa value and random forest gives the highest value (98.23%). It is notice that all algorithms provide relatively higher precision and random forest gives the highest value (98.5%). It can be clearly seen that all algorithms provide relatively higher recall value and random forest gives the highest value (98.5%). It is worth mentioned that all algorithms provide relatively higher specificity value and random forest gives the highest value (99.7%).

For Table 10, RF and KNN provide relatively higher accuracy and RF has the highest accuracy (91.10%). It is noticed that RF and KNN provide relatively higher kappa statistic and RF has the highest kappa value (90.06%). It is worth mentioning that RF and KNN provide relatively higher precision and RF has the highest value (91.20%). It can be clearly seen that RF and KNN provide relatively higher recall and RF has the highest value (91.10%). It is noticed that RF and KNN provide relatively higher specificity and RF has the highest value (99%).

From Table 11, it can be seen that all the algorithms are highly significant for breast cancer and image segmentation dataset. The Random Forest (RF) algorithm has the higher value of F-measure on 3 out of 7 datasets (breast cancer, image, and artificial characters). The LR also gives the higher value of F-measure for vehicle (80.5%) and vertebral column (88.70%). K-NN also provides the highest significant value of F-measure for breast tissue dataset. Bagging provides the highest value of F-measure (53.60%) for CMC dataset. For image and artificial dataset Random Forest (RF) provides highest value of F-measure (98.5% and 91.10%). It is seen that for image segmentation and artificial dataset, the RF provides the highest MCC values are (0.98 and 0.90). For breast cancer, vehicle, and vertebral column dataset, the LR gives the highest MCC values are (0.97, 0.74 and 0.84). For breast tissue dataset, KNN provides the largest value of MCC (0.80) and for CMC dataset all algorithm shows the lowest value of MCC. We also see that the LR gives the best MCC value in 3 out of 7 datasets. It is worth mentioned that for breast cancer dataset, the SVM gives the lesser error rate (0.03) as it provides more perfect prediction and lesser variance in predictions. For vehicle dataset, the LR and NN algorithm provides smallest error rate (0.12). For vertebral column and breast tissue dataset, the LR provides smaller error rate (0.10 and 0.11). For CMC dataset, C4.5 and CART shows the lesser error rate are (0.35). For image dataset, the NN, KNN and CART provides the smaller error

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 03, No. 02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.32.14.32

rate (0.01). For artificial dataset, the KNN and RF performs better as they provide lowest error rate (0.02). It is notice that the Random Forest (RF) algorithm provides lowest root mean square error of 5 (breast cancer (0.15), vehicle (0.28), breast tissue (0.27), image segmentation (0.07) and artificial characters (0.13)) out of 7 datasets. For vertebral column dataset, the LR provides lesser error rate (0.22) than the other algorithms. For CMC dataset, the KNN provides the lowest root mean square error (0.43).

4. CONCLUSION

Twelve different machine learning algorithms were considered: NB, LDA, LR, ANN, SVM, K-NN, HT, DT, C4.5, CART, RF and BB that have been applied on seven datasets. The results show that Random Forest (RF) was found to be the algorithms with most accuracy, precision and Matthew's correlation coefficient (MCC). Other four algorithms: NN, NB, BB and LR were found to be the next accurate after RF accordingly. While kappa statistic and RMSE is another factor. For future research, there is a plan to make hybridization of SVM and ANN to enhance and improve the performance of that type of successful famous machine learning algorithm.

5. REFERENCES

1. Agarap, A. F. M. (2018). On Breast Cancer Detection: An Application of Machine, Learning Algorithms on the Wisconsin Diagnostic Dataset, arXiv:1711.07831v4.
2. Caruana, R., and Niculescu-Mizil, A. (2006). An Empirical Comparison of Supervised Learning Algorithms. Proceedings of the 23rd International Conference on Machine Learning.
3. Helwana et al. (2017). Machine learning techniques for classification of breast tissue, 9th International Conference on Theory and Application of Soft Computing, 120, 402–410.
4. Tekur. A., and Prerna, J. (2018). A Study on Classification Algorithms for Predicting Colon Cancer using Gene Tissue Parameters. IJPAM, 119(18), 2147-2166.
5. Doulah, M.S.U. (2019). A Comparison among Twenty-Seven Normality Tests. Research and Reviews: Journal of Statistics, 8(3), 41-59.
6. Doulah, M.S.U., and Islam M.H. (2019). An Alternative Robust Measure of Outlier Detection in Univariate Data Sets. Research and Reviews: Journal of Statistics, 8(1), 1–11.
7. Doulah, M.S.U., and Islam M.H. (2018). Alternative Robust Methods of Multivariate Outlier Detection. Journal of Mathematical and Statistical Analysis 1(2), 1-9.
8. Doulah, M.S.U. (2018). Alternative Measures of Standard Deviation Coefficient of Variation and Standard Error. International Journal of Statistics and Applications 8(6), 309-315.
9. Doulah, M.S.U. (2021). An Alternative Measures of Moments Skewness Kurtosis and JB Test of Normality. Journal of Statistical Theory and Applications 20(2), 219–227.
10. Padmapriya, B., and Velmurugan, T. (2016). Classification Algorithm Based Analysis of Breast Cancer Data. IJDMTA, 5(1), 43-49.
11. James et al. (2014). An Introduction to Statistical Learning with Applications in R. Springer,
12. Nisbet et al. (2009). Handbook of Statistical Analysis and Data Mining Applications.
13. Senthil et al. (2017). Comparison of Classification Algorithms for Predicting Breast Cancer, IJSRD|, 4(12), 390-394.

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 03, No. 02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.32.14.32

14. Li, M., and Sundararajan, R. (2018). Application of Machine Learning Algorithms on Breast Cancer Dataset, Electrostatics Joint Conference, 1-7.
15. Dana et al. (2018). Diagnostic Accuracy of Different Machine Learning Algorithms for Breast Cancer Risk Calculation: a Meta-Analysis, Asian Pacific Journal of Cancer Prevention, 19(7), 1747-1752.
16. Sadhana et al. (2014). Proportional learning of classifiers using breast cancer datasets. IJCSMC, 3(1), 25-37.
17. Doulah, M.S.U. (2019). Time Series Forecasting: A Comparative Study of VAR ANN and SVM Models, Journal of Statistical and Econometric Methods, 8(3), 21-34.
18. Ravi, I. (2018). Analysis of Classification Algorithms towards Breast Tissue Data Set, IJMTE, 8(12), 650-658.
19. Dietterich, T.G. (2000). An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: BBR, Kluwer-Academic-Publishers, 40, 139–157.
20. Subramani, et al. (2018). Vehicle Identification using Support Vector Machines, IOSRJEN, 18-21.
21. He, H. Garcia, E.A. (2009). Learning from imbalanced data, IEEE Trans. Knowledge Data Eng. 21(9), 1263–1284
22. Ansari et al. (2013). Diagnosis of Vertebral Column Disorders Using Machine Learning Classifiers, Conference: Information Science and Applications (ICISA), 57-89.
23. Huang et al. (2014). Diagnostic Prediction of Vertebral Column Using Rough Set Theory and Neural Network Technique, Information Technology Journal, 13(5), 874-884.
24. Kumar et al. (2009). Guidelines to Select Machine Learning Scheme for Classification of Biomedical Datasets, 128-139.
25. Orsu et al (2013). Performance Analysis and Evaluation of Different Data Mining Algorithms used for Cancer Classification. IJARAI, 2(2), 67-89.
26. Han J, Kamber M. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, San Francisco. 2006; 443-540.
27. Everitt BS. Cluster analysis. Edward Arnold, London, 1993.
28. Jakkula, V. (2011). Tutorial on Support Vector Machine (SVM). Retrieved August 15, 2011, from http://eecs.wsu.edu/~vjakkula/SVMTutorial.doc
29. Khanna et al. (2015). Comparative Study of Classification Techniques (SVM, Logistic Regression and Neural Networks) to Predict the Prevalence of Heart Disease. IJMLC, 5(5), 414-419.
30. Doulah, M.S.U. (2019). Application of Machine Learning Algorithms in Bioinformatics, Bioinformatics and Proteomics Open Access Journal, 3(1), 1-11.
31. Berardi et al. (2004). A principled Approach for Building and Evaluating Neural Network Classification Models. Decision Support Systems, 233-246.
32. Breiman et al. (1984). Classification and Regression Trees, Wadsworth, Belmont.
33. Dogan, N., and Tanrikulu, Z. (2010). A Comparative Framework for Evaluating Classification Algorithms. Proceedings of the World Congress of Engineering.
34. Duda et al. (2001). Pattern Classification, Wiley-New York.
35. Srimani et al. (2015). Performance analysis of Hoeffding trees in data streams by using massive online analysis framework. Int. J. Data Mining, Modelling and Management, 7(4), 293-313.

Journal of Artificial Intelligence, Machine Learning and Neural Network
ISSN: 2799-1172
Vol: 03, No. 02, Feb-Mar 2023
http://journal.hmjournals.com/index.php/JAIMLNN
DOI: https://doi.org/10.55529/jaimlnn.32.14.32

36. Wang et al. (2016). An Algorithm Model for Incremental detection of Spam Reviews. International Journal of Modeling and Optimization, 6(1), 45-48.
37. Tharwat A. (2021). Classification assessment methods, Applied Computing and Informatics, 17(1), 39-56.
38. Yu-Wei, and Chiu, D. (2015). Machine learning with R cookbook, Published by Packt Publishing Ltd. Birmingham B3 2PB, UK
39. https://archive.ics.uci.edu/ml/datasets/
40. Oladugba et al. (2018). On the Comparison of Methods of Estimating Missing Values in Rectangular Lattice Designs, Journal of Mathematics and Statistics, 14, 201-208.
41. Doulah, M.S.U. (2022). Bioinformatics Biostatistics and Biometrics: A Statistical Journey. Annal Biostat and Biomed Appli. 4(3), 1-3.