# Adam optimized Logistic Regression Model to improve heart disease diagnosis

**Shiwani Gupta[1], R. R. Sedamkar[2]**

[1]Research Scholar,
*TCET Mumbai,*
*shiwani.gupta@thakureducation.org*

[2]Professor
*TCET Mumbai,*
*rr.sedamkar@thakureducation.org*

***Abstract*: *Machine Learning models have been experimented a lot by researchers for heart disease prediction. The power of a simplistic interpretable model as Logistic Regression can be enhanced comparable to an ensemble. This experimentation has been done on Framingham heart disease data that is relatable by medical practitioners and the features are interpretable. Logistic Regression utilizes sigmoid function and log loss is computed for the loss function. To optimize the same, gradient descent is used which uses static learning rate and converges a convex cost function. The Learning curve for Logistic Regression classifier with optimal threshold through PRCurve was found to be neither biased nor had variance. It even failed to converge. Thus, in order to improve performance, ADAptive Moment optimizer is used which accelerates training by smoothing learning and uses an adaptive learning rate per parameter. This stochastic optimization of a non-convex loss function has improved weighted Fscore. The novel algorithm has also been tested on benchmarked Statlog, Cleveland and CVD datasets with improvement in other metrics as well. The results have been statistically validated through McNemar's non parametric test and 5X2 CV paired t-test.***

*Keywords*: **Logistic Regression, Adam, Framingham, McNemar**

**(Article history: Received: 6th April 2021 and accepted 16th December 2021)**

## I. INTRODUCTION

Heart failure (HF) is a major public health issue, with a prevalence of over 5.8 million in U.S. and over 23 million worldwide and rising. The lifetime risk of developing HF is one in five [1].

Selection of ML algorithm should be based on the question - how large the population is, how many cases exist, how balanced the dataset is, how many available variables are there, the clinical outcome is binary or not, etc. [2].

This section reviews methodology and research issues in various literature published particularly on use of optimization to enhance the performance of Logistic Regression for prediction of heart disease and choice of statistical hypothesis tests to check statistical significance of results.

Discriminative classifiers as Logistic Regression are better than generative ones as Naïve Bayes [3]. Logistic Regression model is better than Neural Network on Cleveland heart disease dataset [4]. Logistic Regression has better accuracy as compared to other classifiers and complex relationships between dependent and independent variables are identified easily [5]. Logistic Regression is second good classifier which improves with feature selection and hyperparameter tuning [6]. Logistic Regression is the least accurate w.r.t. Support Vector Machine and Artificial Neural Network on Z-Alizadehsani dataset [7]. For imbalanced data, Logistic Regression performs better in terms of mean Area Under the Curve [8]. Logistic Regression was seen as the best classifier providing 85% accuracy and 87% Fscore on Statlog dataset [9]. Logistic Regression on Statlog gave 85% accuracy, 86% precision, 80% recall and 91.82% AUC [10]. Integrated classifier (Random Forest) gave 78% accuracy, 75% sensitivity and 80% specificity on test set for Framingham heart disease data [11].

Gradient Descent optimisation is a black box optimiser with variants as batch, stochastic and minibatch. Other optimization algorithms as Momentum, Nestorov, Adagrad, Adadelta, RMSProp, Adam, Adamax, NAdam have been compared [12]. Adam, an algorithm for first order gradient based optimization of stochastic objective function is easy to implement, computationally efficient and has less memory requirements. It is appropriate for noisy gradients [13]. LogitBoost provides convex optimization [14].

[15] has used Shapiro Wilks normality test, Welch's t-test, Mann Whitney Wilcoxon test, Pearson Chi2 independence test and Fisher's exact test to identify relation between features. Parametric tests as Pearson and Fisher score and non parametric tests as Kendall's Tau and Spearman Rho have been used for feature selection on Alizadehsani heart disease data [16]. Genetic Algorithm and Information Gain Models are suitable for low sized data while tree based models are suitable for high dimensional data [17]. [18] states that repeated CV provides bias and wastes computational resources though reducing the variance. McNemar test is the only test with acceptable Type I error for algorithms that can be executed only once. For small datasets, 5X2 CV test is recommended [19].

Logistic Regression works well for binary outcome and outliers need to be removed. Being a linear algorithm, it assumes linear relationship between predictor and explanatory variable. The algorithm might fail to converge with sparse or multicollinear data. It is basically a sigmoid function with an S shaped curve that maps a real valued number to a value between 0 and 1. It is a probabilistic model which compares the prediction probability with threshold which is by default 0.5. Since Logistic Regression is derived from Linear Regression as shown in Eq. 1 below, the coefficients β0, β1 need to be estimated by Maximum Likelihood Estimation (MLE) method.

$$h_\theta(X) = 1/(1+e^{(-(\beta 0+ \beta 1X))}) \quad (1)$$

The cost function of Logistic Regression as shown in Eq. 2 below is minimized by applying Gradient Descent optimization algorithm which takes partial derivative of cost w.r.t. parameters and updates these parameters for each iteration with a selected learning rate α as shown in Eq. 3 below until the gradient has converged.

$$J(\theta) = -(1/m) * \Sigma\ [y^{(i)} * \log(h_\theta(x^{(i)})) + (1-y^{(i)}) * \log(1-h_\theta(x^{(i)}))] \quad (2)$$

where m is the no. of samples and log is taken to ease the derivative.

Repeat {

$$\Theta_j := \theta_j - \alpha * \Sigma\ (h_\Theta(x^{(i)})-y^{(i)}) * x_j^{(i)} \quad (3)$$

#simultaneously update all Θj, i = 1 to m

}

here superscript denotes iterations

This paper introduces a novel method to optimize the loss function of Logistic Regression classifier. The major contributions are enlisted as follows:

- Python implementation of regularized Logistic Regression classifier uses stochastic average gradient descent (sag) optimizer that shows fluctuations. Further, it doesn't converge during execution.
- Adaptive moment estimation method to optimize the loss function of Logistic Regression classifier has been used since the loss function equation is non convex and Gradient Descent works well with Linear functions leading to suboptimal local minima.
- Gradient Descent takes static learning rate for all parameters whereas one should perform larger updates for sparse features. Adaptive moment estimation is a non-convex optimizer that provides bias correction. Momentum accelerates training by smoothing learning and thus overcomes the oscillations of noisy gradient leading to faster convergence and better generalization and it also considers separate learning rate per parameter.

Rest of the paper is organized as follows: Section 2 explains materials and methods. Section 3 explains the results and discussion and Section 4 concludes the work.

## II. MATERIALS AND METHODS

This section states the experimental work in detail including description of methodology, proposed algorithm, datasets, evaluation metrics and statistical tools.

### A. METHODOLOGY

Fig. 1 describes the general architecture used followed by the generalized algorithm designed for experimentation on Framingham Heart Disease Dataset [20,21]:

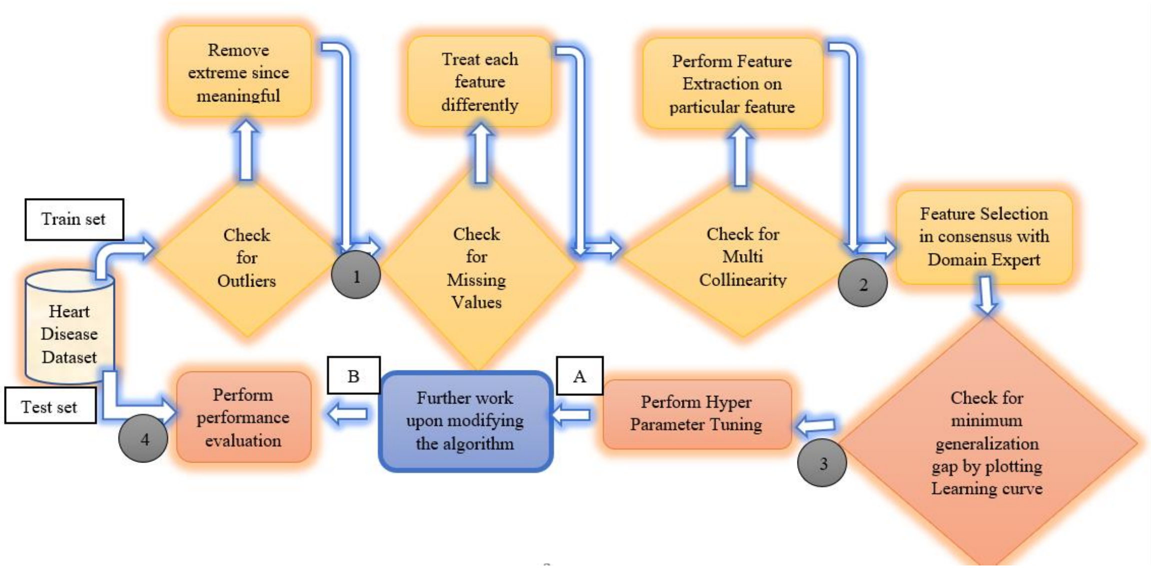


Fig. 1. General Architecture (CFSnRGG Model with embedded AoLoR Model)

B. ALGORITHMIC DESCRIPTION OF CONSENSUS OF FEATURE SELECTION AND REDUCED GENERALIZATON GAP MODEL [27]

1. Perform Data Preprocessing
   a. Check for skewness, kurtosis and outliers, treat extreme outliers only since features are meaningful.
   b. Data has missing values, treat each feature differently.
2. Exploratory Data Analysis
   a. Check multi collinearity with VIF, treat with PCA.
   b. Check Correlation with heatmap, perform feature selection in consensus with domain expert.
3. Model Building
   a. Build Logistic Regressor with optimal threshold through PRCurve.
   b. Plot Learning curve to check whether model is suffering from bias or variance
   c. Perform hyper parameter tuning through Validation Curve and RandomizedSearchCV to reduce bias and generalization gap.
   d. Perform RobustScaler since model has meaningful outliers.
   e. Train model with train set size of min generalization gap derived from Learning Curve in comparison to static split of 70:30.
4. Model Evaluation
   a. Data is imbalanced, chose appropriate performance metric as weighted f score and model evaluation technique as Stratified k fold CV.

C. AoLoR MODEL

Gradient Descent is one of the most popular algorithms to perform optimization which assumes the terrain to be convex. It minimizes the objective function by updating the parameters as shown in Eq. 4 in opposite direction of the gradient of objective function w.r.t. the parameters. The learning rate, η determines the size of steps we take to reach a minima. Choice of an appropriate learning rate is difficult. Secondly, same learning rate applies to all parameter updates, whereas in case of sparse data, one should perform a larger update for rarely occurring features. Moreover, minimizing a highly non convex error function leads to suboptimal local minima.

$$\Theta = \theta - \eta * \nabla_\theta J(\theta)$$

where $\nabla_\theta J(\theta)$ is partial order derivative of $J(\theta)$ w.r.t. θ

Stochastic Gradient Descent (S.G.D) computes updates for each individual sample. Momentum is introduced to accelerate gradient descent in order to dampen oscillations and speed convergence as illustrated in Eq. 5 and 6 below. $v_{t-1}$ is forward looking gradient.

$$v_t = \gamma * v_{t-1} + \eta * \nabla_\theta J(\theta) \quad (5)$$

$$\Theta = \theta - v_t \quad (6)$$

S.G.D. with momentum above calculates previous gradient and makes a jump then calculates current gradient from same point and makes another jump. The momentum term is usually set to 0.9. 1st small blue arrow in Fig. 2 below demonstrates momentum current gradient, next arrow represents updated gradient. Big red arrow represents NAG big jump in direction of previous gradient. Black arrow represents correction. Next arrow represents final jump from point of correction.

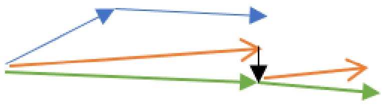

Fig. 2. Gradient Update

Since our data is sparse i.e. imbalanced, the learning rate is adapted for parameters with frequently and not frequently occurring features as shown in Eq. 7, update rule – Eq. 8 and modified rule Eq. 9 below:

$$g_{t,i} = \nabla_{\theta} J(\theta_{t,i}) \quad (7)$$

$$\theta_{t+1,i} = \theta_{t,i} - \eta * g_{t,i} \quad (8)$$

$$\theta_{t+1,i} = \theta_{t,i} - (\eta / \sqrt{(G_{t,ii}+\epsilon)}) * g_{t,i} \quad (9)$$

where $G_{t,ii}$ is the diagonal matrix where each diagonal element ii is sum of square of gradients w.r.t. $\theta_i$ upto time step t, $\epsilon$ term avoids division by zero. It eliminates the need of manually tuning the learning rate, which eventually becomes zero and the learning stops.

Since the learning rate diminishes monotonically because the denominator is cumulative sum of squared gradients, we divide it by average of squared gradients. A bias correction and momentum is added to compute the estimates of first moment (weighted mean $m_t$) and second moment (uncentered variance $v_t$). The adam update rule is mentioned in Eq. 10 below. The default values for β1, β2 and $\epsilon$ is 0.9, 0.999 and $10^{-8}$ respectively. Thus high variance gradient in different direction take smaller steps.

$$\theta_{t+1} = \theta_t - (\eta / \sqrt{(v_t+\epsilon)}) * m_t \quad (10)$$

where $m_t = m_t / (1 - \beta_1 t)$ and $v_t = v_t / (1 - \beta_2 t)$ are bias corrected forms, higher the β, less we update current updates which give smoother movement.

Thus AoLoR model is demonstrated in Fig. 3 below:

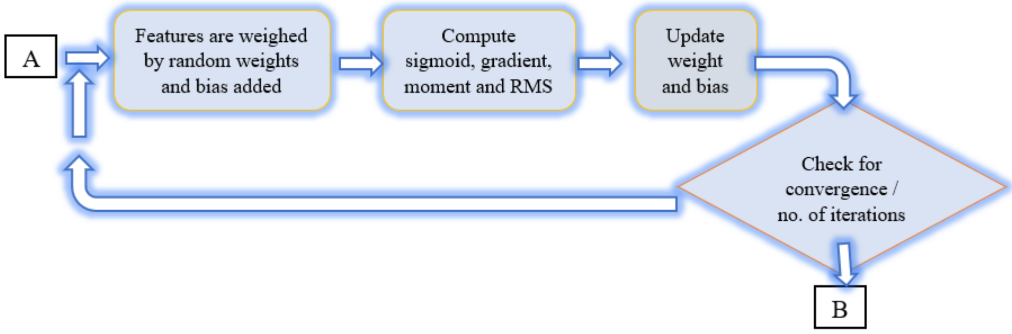


Fig. 3. AoLoR Model Diagram

## D. PSEUDOCODE FOR AoLoR ALGORITHM

| | **Algorithm** |
|---|---|
| 1 | **Inputs:** |
| 2 | Features and Label |
| 3 | **Output:** |
| 4 | Class prediction |
| 5 | **Begin** |
| 6 | Initialize learning_rate, n_iter, beta_1, beta_2, epsilon, moment for weights, rms prop for weights, moment for bias, rms prop for bias |
| **7** | Derive rows and columns from dataset |
| 8 | Initialize weights and bias |
| 9 | **For** n_iter |
| 10 | Take dot product of X matrix with weights and add bias |
| 11 | Compute sigmoid |
| 12 | Compute gradient for weights and biases |
| 13 | Compute moments and RMS for weights |
| 14 | Compute moments and RMS for bias |

```
Update weight and bias
End For
Return prediction
End
```

Fig. 4. Pseudocode of Proposed Model

## E. DATASET DESCRIPTION

Framingham [22] Heart Disease data predicting ten year Coronary Heart disease (CHD) with 4,225 instances and 15 features are preferred for experimentation due to availability of more data and less percentage of missing values. Any disease dataset has been found to be imbalanced since number of instances in healthy class will be more as compared to diseased class. Statlog [23,24] and Cleveland [23] heart disease datasets both have 270 and 303 instances respectively with same 12 attributes. CVD dataset has 70000 instances with 12 attributes [25].

## F. EVALUATION METRIC

Accuracy is the most intuitive performance measure

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \quad (11)$$

where TP is correctly predicted positive value, TN is correctly predicted negative value, FP is falsely predicting as positive, FN is falsely predicting as negative

Precision (P) is the ratio of correctly predicted positive observations to total predicted positive observations

$$\text{Precision}=TP/(TP+FP) \quad (12)$$

Recall (R)/ Sensitivity is the ratio of correctly predicted positive observations to all actual positive observations

$$\text{Recall}=TP/(TP+FN) \quad (13)$$

F1 score is weighted average of Precision and Recall. It is especially useful for imbalanced data.

$$\text{F1score}=2*P*R/(P+R) \quad (14)$$

Weighted Fscore is the F1score of each label weighted by support. It emphasizes the importance of some samples w.r.t. the others.

Area Under the Curve (AUC) measures the 2-D area underneath Receiver Operating Characteristics (ROC) curve. This graph shows performance of classification model at all classification threshol It is a plot of True Positive Rate (TPR) vs Fa Positive Rate (FPR).

## G. STATISTICAL TOOLS

Models are being evaluated using resampling method. The difference between mean of two model could be due to statistical chance. Thus, taking null hypothesis $H_0$ as the assumption that both models are same or none of two models perform better than the other and $H_1$ as the assumption that perfromances of two models are not equal. Same is shown in Eq. 15 below. In order to improve confidence in interpretation and representation of results during model selection i.e. the trust in estimated skill of each model and thus strengthening the claim, statistical hypothesis tests have been used.

$$H_0 : p_b = p_c;\ H_1 : p_b \neq p_c \quad (15)$$

Since the distribution of estimates are not Gaussian in nature, McNemar's nonparametric statistical test is performed on paired nominal data through a 2X2 contingency table as in Table 1 below, which checks marginal homogeneity of 2 dichotomous variables and returns the test statistic as per Eq. 16. below and p-value. It is used when data of 2 groups is coming from same participants. If $p>\alpha$, the significance level; the null hypothesis is accepted which means the two models are not statistically significant.

TABLE 1: CONTINGENCY TABLE

| | **Model 2 correct** | **Model 2 wrong** |
|---|---|---|
| **Model 1 correct** | A | b |
| **Model 1 wrong** | C | d |

$$\chi^2 = (b-c)^2/(b+c) \quad (16)$$

where b=y/n and c=n/y are discordant cells

The assumptions of McNemar's test are being followed stating the suitability for choice of test as follows:

1. Presence of one nominal variable with two categories (i.e. dichotomous variables) as predicted 'CHD' and one independent variable with two connected groups as actual 'CHD'.
2. The two groups in your dependent variable must be mutually exclusive. Thus the participants are not appearing in more than one group. The

sample drawn is random since Cross Validation has been used.

In 5X2 CV paired t-test, we split data into 2 parts: training and test and repeat splitting 5 times. In each of 5 iterations, we fit both models on training split and evaluate their performance on test split. Post rotating the training and test set, we compute the performance again. The mean and variance of differences is reported. Null Hypothesis assumes that both models have equal performance. If $p < \alpha$, the significance level, we reject null hypothesis and accept that there is significant difference in two models. Score differences in 5X2 fold CV test are computed through t statistic in Eq. 17. below:

$$t = p_1^{(1)} / (\sqrt{(1/5(\sum_{i=1}^{5} S_i^2)}))$$

where $p_1^{(1)}$ is the classifiers' scores difference for the first fold of the first iteration, $s_i^2$ is the estimated variance of the score difference for $i^{th}$ iteration. This variance computes as $(p_i^{(1)} - \bar{p}_i)^2 + (p_i^{(2)} - \bar{p}_i)^2$, where $p_i^{(j)}$ is the classifiers' scores difference for the $i^{th}$ iteration and fold j, and $\bar{p}_i = (p_i^{(1)} + p_i^{(2)})/2)$

## III. RESULT AND DISCUSSION

Result has been compared for CFSnRGG Model i.e. Balanced Regularized (L2) Logistic Regression with optimal threshold of 0.20 and train test split obtained from learning curve with minimal generalization gap i.e. 1500 training instances and hypertuned value of C=2.07 and Base Model i.e. Regularized Logistic Regression with static train test split of 70:30 i.e. 2958 training instances for Framingham Heart disease data. Same is presented in Table 2 below. The 5X2 CV t test statistic shows that there is acceptable difference in the two models statistically as $\bar{t} < -2$. 78% accuracy was reported by [11] on Framingham Heart Disease data which is lesser than proposed model.

TABLE 2: CFSNRGG MODEL VS BASE MODEL ON FRAMINGHAM DATA

| | **Weighted F score** | **Precision** | **Accuracy** | **5X2 CV** |
|---|---|---|---|---|
| Base Model | 70.7+/-2.6 | 26+/-3.2 | 66.1/-3 | **t test statistic=** -3.06 |
| Proposed Model | **80.0+/-1.1** | **68+/-25** | **85.4+/-0.5** | |

But CFSnRGG Model had no scope of enhancing performance further as seen from Learning curve in Fig. 5. below. Hence we designed AoLoR Model.


Fig. 5. Learning Curve for Framingham demonstrates neither overfitted nor underfitted model

(17) Results of experimentation on Framingham demonstrate enhanced performance with proposed AoLoR Model i.e. Logistic Regression with Adam optimizer w.r.t. Base Model i.e. Logistic Regression with gradient descent optimizer in terms of multiple evaluation metrics. McNemar Test pvalue = 1 < statistic = 3; hence we reject null hypothesis, thus both models are statistically significant. Moreover 5X2 CV t test statistic shows that there is acceptable difference in the two models statistically as $\bar{t} < -2$. Drawback of Python implementation of Logistic Regression classifier is that it provides a warning "Total no of iterations reached limit". Fig. 6 demonstrates enhanced performance of designed novel model w.r.t. Python variant.

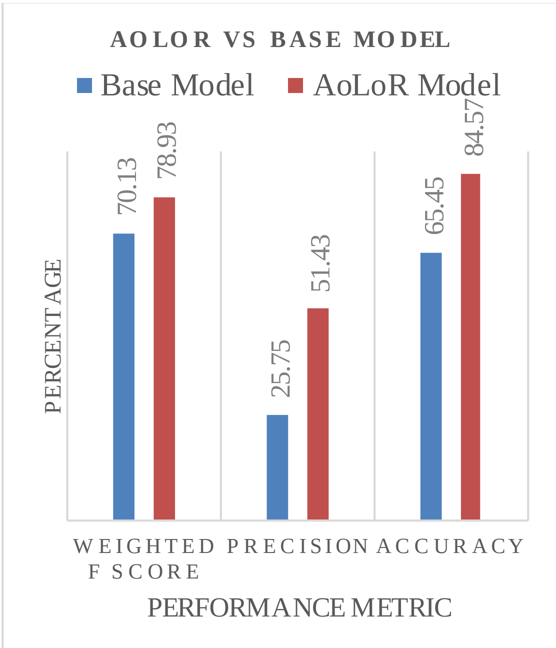


Fig. 6. AoLoR Model vs Base Model on Framingham Data

Experimentations on Framingham heart disease data were not found in Literature reviewed. Hence, to validate above experimentation, we applied

CFSnRGG Model onto benchmarked Statlog Heart Disease Data [9,10] where Logistic Regression classifier was used. Results have been demonstrated in Fig. 7. below. Enhancement in performance for several metric can be seen.

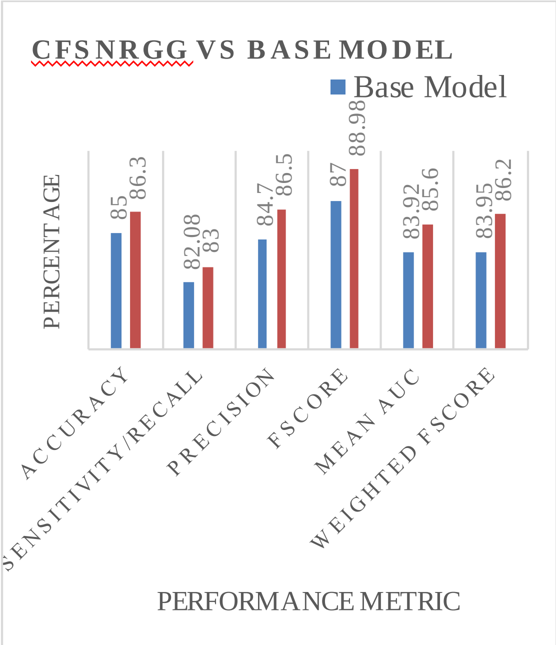


Fig. 7. CFSnRGG Model vs Base Model on Framingham Data

Results of experimentation on Statlog Heart Disease Data with AoLoR Model i.e. Logistic Regression with adam optimizer onto 90:10 split since data is small w.r.t. CFSnRGG Model i.e. balanced regularized (L2) Logistic Regression with optimal threshold and hypertuned utilizing gradient descent optimizer in terms of multiple evaluation metrics are demonstrated in Fig. 8. below. McNemar Test pvalue = 0.23975 < statistic = 0.5 hence we reject null hypothesis, thus both models are statistically significant. Sensitivity achieved is more than 85.8% in [24]. Decision Tree and Support Vector Machine gave 72.5% and 68.75% sensitivity and 75.79% and 75.26% accuracy respectively on Statlog heart disease data [26].

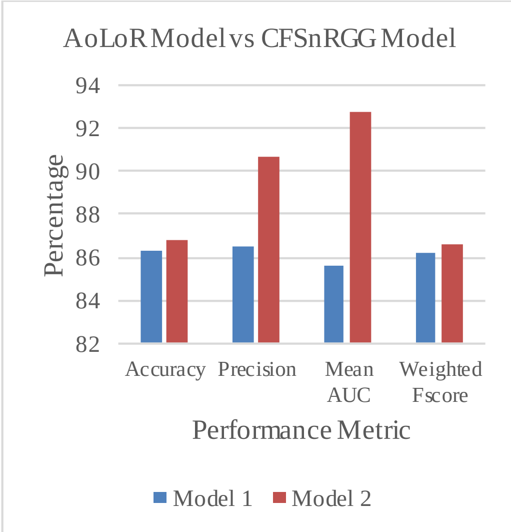


Fig. 8. AoLoR Model vs CFSnRGG Model on Statlog Data

Since Learning Curve as shown in Fig. 9. below for Statlog shows need of more data for convergence of train and validation performance, We merged Cleveland and Statlog datasets with 525 instances and performed experimentation for enhanced performance.

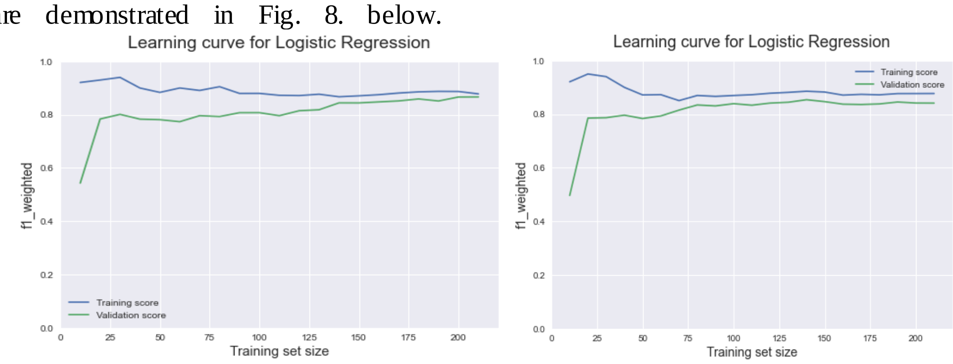


Fig. 9. Learning Curve for Statlog and Cleveland demonstrate need of more training data

Comparing AoLoR Model and CFSnRGG Model on Cleveland and Statlog combined dataset, we can see AoLoR Model outperforms CFSnRGG Model in Fig. 10. below.


Fig. 10. AoLoR vs CFSnRGG Model on combined Statlog and Cleveland Data

McNemar Test pvalue = 0.23975 < statistic = 0.5 hence we reject null hypothesis, thus both models are statistically significant.

Experimentation was also done on CVD heart disease data which is of high volume having 70000 instances and 12 features. Results demonstrate enhanced weighted Fscore and Accuracy in Fig. 11 below.

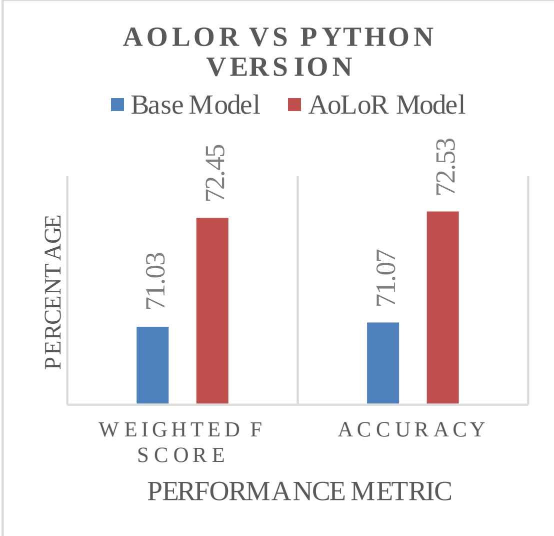


Fig. 11. AoLoR vs Base Model on CVD Data

McNemar Test pvalue = 0.00005 < statistic = 15.197 hence we reject null hypothesis, thus both models are statistically different with 99.99% confidence. $\bar{t}$ = -21.28.

Thus stating comparative analysis with contemporary researchers in Table 3. below:

CFSnRGG Model designed gave higher accuracy by 7.4% and 1.3% respectively and Fscore by 1.98% wrt the referenced literature of 2018 and 2020 on Framingham and Statlog Heart Disease data. AoLoR Model designed achieved higher accuracy by 1.4% and 11.2% respectively on Statlog data wrt referenced literatures of 2016 and 2020. AoLoR Model when applied on combined Statlog and Cleveland, due to increased no. of instances received better performance than Cleveland reported in referenced literature of 2016 and 2020 wrt accuracy by 13.75%, Sensitivity by 9.5% and RoCAuC by 4%.

TABLE 3: CFSnRGG MODEL AND AoLoR MODEL COMPARED WITH BENCHMARK RESULTS

| Dataset | Model | Result | Methodology | Benchmark Result | Reference |
|---|---|---|---|---|---|
| Framingham Heart Disease | CFSnRGG MODEL | 85.4% accuracy | Python Logistic Regression | 78% accuracy | PLoS ONE 2018 [18] |
| Statlog Heart Disease Data | | 86.3% accuracy and 88.98% Fscore | | 85% accuracy and 87% Fscore | Hindawi BioMed Research International 2020 [19] |
| | AoLoR MODEL | 87% accuracy | Least Squares Twin Support SVM | 85.59% accuracy | IEEEAccess 2020 [20] |
| | | | DT, SVM | 75.79% and 75.26% accuracy | Computational Intelligence 2016 [21] |
| Statlog and | | 82.5% Sensitivity | | 72.5% and 68.75% | |

| | | | | | |
|---|---|---|---|---|---|
| Cleveland combined dataset | | | | sensitivity | |
| | | 84% accuracy, .925 RoCAuC, 82.5% Sensitivity | MFFSA and AFSA | 82.9% accuracy, 0.885 RoCAuC, 75% sensitivity | Elsevier ScienceDirect Computers and Electrical Engineering 2020 [24] |
| | | 82.5% Sensitivity | FCMIM-SVM | 75% Sensitivity | IEEEAccess 2020 [25] |

## IV. CONCLUSION

Thus, the chosen problem of heart disease diagnosis has been experimented through several stages of Machine Learning cycle i.e. EDA, Feature selection, extraction, model selection, model evaluation, etc. In order to create a generalized model, minimum generalization gap is identified. The performance of Logistic Regression classifier has been found to enhance for several metrics utilizing CFSnRGG Model. But to enhance the performance further, AoLoR Model has been designed so that convergence happens. The same mathematical model has been tested on benchmark datasets and statistical comparison made of two models. The results infer that designed models show statistical significance. Same model shall be applicable to other critical disease data as Cancer, Pima Indian Diabetes etc. since medical data is sparse.

## ACKNOWLEDGMENT

We are extremely thankful to my reviewer Dr. S. K. Shinde and Dr. Archana Patankar for their valuable suggestions. I thank the students who helped me with the code. I even thank industry persons who have validated my claims.

## REFERENCES

[1] A. Bui, T. Horwich, G. Fonarow "Epidemiology and risk profile of heart failure" Nat Rev Cardiol 8(1):30-41, 2011. doi: 10.1038/nrcardio.2010.165.

[2] C. Krittanawong, H. U. H. Virk, S. Bangalore, Z. Wang, K. W. Johnson, R. Pinotti, H. J. Zang, S. Kaplin, B. Narsimhan, T. Kitai, U. Baber, J. L Halperin., W. H. W. Tang "Machine Learning prediction in cardiovascular diseases: a meta analysis" SCIENTIFIC REPORTs 10:16057, 2020. https://doi.org/10.1038/s41598-020-72685-1.

[3] A. Y. Ng and M. I. Jordan "On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes" Advances in neural information processing systems pp. 841–848, 2002.https://papers.nips.cc/paper/2001/file/7b7a53e239400a13bd6be6c91c4f6c4e-Paper.pdf

[4] J. Premsmith and H. Ketmaneechairat "A Predictive Model for Heart Disease Detection Using Data Mining Techniques" Journal of Advances in Information Technology, (12) 1, 14-20 2021. doi: 10.12720/jait.12.1.14-20.

[5] C. C. Nuria, L. C. S. José, A. G. P. Juan, M. G. P. José and L. P. L. María "Machine Learning Applied to Diagnosis of Human Diseases: A Systematic Review" Appl. Sci. 10, 5135, 2020. doi:10.3390/app10155135.

[6] J. Ping, A. U. Haq, S. U. Din, J. Khan, A. Khan and A. Saboor "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare", 2020. Digital Object Identifier 10.1109/ACCESS.2020.3001149

[7] I. C. Dipto, T. Islam, H. M. M. Rahman and M. A. Rahman "Comparison of Different Machine Learning Algorithms for the Prediction of Coronary Artery Disease". Journal of Data Analysis and Information Processing, 8, 41-68, 2020. https://doi.org/10.4236/jdaip.2020.82003

[8] X. Zheng "SMOTE Variants for Imbalanced Binary Classification: Heart Disease Prediction" UCLA Electronic Theses and Dissertations 2020, https://escholarship.org/uc/item/99x0w9w0.

[9] A. T. Bayu, I. Sun and L. Seungchul "Improving an Intelligent Detection System for Coronary Heart Disease Using a Two-Tier Classifier Ensemble" Hindawi BioMed Research International Volume Article ID 9816142, 10 pages, 2020. https://doi.org/10.1155/2020/9816142

[10] Q. Zhenya and Z. Zuoru "A hybrid cost-sensitive ensemble for heart disease prediction" BMC Medical informatics and Decision making 2020. DOI: 10.21203/rs.2.22946/v3

[11] M. V. Dogan, I. M. Grumbach, J. J. Michaelson, R. A. Philibert "Integrated genetic and epigenetic prediction of coronary heart disease in the Framingham Heart Study". PLoS ONE 13(1): e0190549, 2018. https://doi.org/10.1371/journal.pone.0190549

[12] S. H. Haji, A. M. Abdulazeez "Comparison of optimisation techniques based on gradient descent algorithm: A review" PJAEE 18 (4) 2021. https://archives.palarch.nl/index.php/jae/article/view/6705

[13] D. P. Kingma "Adam: A method for stochastic optimization", 2017 arXiv:1412.6980v9 [cs.LG]

[14] C. Cortes, V. Vapnik "Support-vector networks" Mach. Learn. 20, 273–297, 1995. https://doi.org/10.1007/BF00994018

[15] F. Babič, J. Olejár, Z. Vantová, J. Paralič "Predictive and Descriptive Analysis for Heart Disease Diagnosis" Proceedings of the Federated Conference on Computer Science and Information Systems pp. 155–163 ACSIS, Vol. 11, 2017. DOI: 10.15439/2017F219 ISSN 2300-5963

[16] S. Gupta, R. R. Sedamkar "Feature Selection to reduce dimensionality of heart disease dataset without compromising accuracy", International Journal of Computer Trends and Technology (IJCTT) 67, 6, 2019. https://doi.org/10.14445/22312803/IJCTT-V67I6P109

[17] C. H. Liu, C. F. Tsai, K. L. Sue and M. W. Huang "The Feature Selection Effect on Missing Value Imputation of Medical Datasets", Appl. Sci. 10, 2344 2020. doi:10.3390/app10072344 www.mdpi.com/journal/applsci

[18] G. Vanwinckelen and H. Blockeel "On estimating model accuracy with repeated cross validation". Proceedings of BeneLearn and PMLS 2012. https://lirias.kuleuven.be/1655861?limo=0

[19] T. G. Dietterich "Approximate Statistical tests for comparing Supervised Classification algorithms" Neural Computation 10 (): 1895–1923, 1998. https://doi.org/10.1162/089976698300017197.

[20] A. G. E. Hoda, S. Liyakathunisa "An Automatic Early Risk Classification of Hard Coronary Heart Diseases using Framingham Scoring Model", ICC Cambridge, United Kingdom © 2017 ACM. ISBN 978-1-4503-4774-7/17/03$15.00 DOI: http://dx.doi.org/10.1145/3018896.3036384

[21] S. Gupta, R. R. Sedamkar Springer Nature Book: Machine Learning with HealthCare Perspective, Chapter: Machine Learning for Healthcare: Introduction, Learning and Analytics in Intelligent Systems. 13, 2020. https://doi.org/10.1007/978-3-030-40850-3

[22] S. Gupta, R. R. Sedamkar "Genetic Algorithm for Feature Selection and Parameter Optimization to Enhance Learning on Framingham Heart Disease Dataset" In: Balas V.E., Semwal V.B., Khandare A., Patil M. (eds) Intelligent Computing and Networking. Lecture Notes in Networks and Systems, 146, 2021. Springer, Singapore. https://doi.org/10.1007/978-981-15-7421-4_2.

[23] S. Gupta, R. R. Sedamkar "Apply Machine Learning for healthcare to enhance performance and identify informative features" Proceedings of IEEE INDIACom; 6th International Conference on "Computing for Sustainable Global Development, BVICAM, New Delhi (INDIA) 368-372, 2019. https://ieeexplore.ieee.org/document/8991386

[24] J. Wang, C. Liu, L. Li, W. Li, L. Yao, H. Li and H. Zhang "A Stacking-Based Model for Non-Invasive Detection of Coronary Heart Disease IEEEAccess Vol 8, 2020. Digital Object Identifier 10.1109/ACCESS.2020.2975377

[25] https://ieee-dataport.org/documents/cardiovascular- disease-dataset

[26] S. Bashir, U. Qamar, F. H. Khan "A multicriteria weighted vote-based classifier ensemble for heart disease prediction". Computational Intelligence 32(4), 615–645, 2016. https://doi.org/10.1111/coin.12070

[27] S. Gupta, R. R. Sedamkar, "Consensus of Feature Selection methods and Reduced Generalization Gap Model to improve Diagnosis of Heart Disease", Journal of Scientific Research, Vol 3, Issue 13, 1 Sep 2021. UGC Care.

AUTHOR PROFILE

**Shiwani Gupta**

She is currently working as an Assistant Professor in Computer Engineering Department at Thakur College of Engineering and Technology, Mumbai, India. She is pursuing Ph.D. in Technology from University of Mumbai. She holds M.Tech and B.Tech degree in Computer Science and Engineering in 2010 and 2003 respectively. She has over 18 years of experience teaching B.Tech, B.E. MCA and M.Tech courses. Her major area of interest include Artificial Intelligence, Machine Learning and Algorithms. She has around 75 research publications in various international/national journals/conferences.

**R. R. Sedamkar**

He is currently working as a Professor in Computer Engineering Department and Research Centre of Ph.D. Programs at Thakur College of Engineering and Technology, Mumbai, India. He has 7 research scholar working under his guidance currently. He has over 29 years of teaching experience. His area of interest include Networking. He has over 50 research publications in various international/national journals/conferences.