International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Improved Sensitivity and Reliability in Heart Disease Survival Prediction: A Stacking SVM and Logistic Regression Approach

Vijayalakshmi Sarraju[1], Jaya Pal[2], Supreeti Kamilya[3]

Submitted:14/03/2024 **Revised**: 29/04/2024 **Accepted**: 06/05/2024

Abstract: Predicting heart disease survival remains challenging in clinical data analysis due to the complexity and variability of the data sets. This research examines the performance of support vector machine (SVM) and logistic regression (LR) classifiers on a stroke prediction dataset to enhance sensitivity and reliability in heart disease prediction in varying sample sizes. Our analysis shows that SVM is consistently high in precision, specificity, and accuracy, while LR is variable, with a marked drop in sensitivity with increasing sample size. We propose a stacked ensemble model by integrating the strengths of SVM and LR. The stacked ensemble performs best in achieving the highest sensitivity of 0.97, specificity of 0.95, and F1-score of 0.96 in the largest sample size. This method significantly improves prediction accuracy and reliability, which makes it very applicable to early detection and effective management of heart disease.

Keywords: *Support Vector Machine (SVM), Sample size, cardiovascular disease (CVD), Logistic Regression Performance metric, Stacked Ensemble Learning.*

1. Introduction

Support vector machines (SVMs), LR, and entropy approaches have recently been used in medical diagnostics. Because they can handle various high-dimensional information, these approaches are promising for diagnosing, prognosis, and predicting disease treatment. The efficacy of SVM and LR medical diagnostic models depends on several aspects [1]. The influence of sample size on machine learning models affects its capacity to identify patterns and generalize to unknowns. The sample size drastically affects the performance of the model in varied medical datasets [2]. The influence of sample size on the performance of the medical diagnostic model is underexplored despite its importance.

This research examines the impact of sample size on SVM and LR models in diagnosing cardiovascular diseases, exploring their relationship with metrics such as accuracy, specificity, sensitivity, precision, and F1-score. By analyzing performance across varying sample sizes, the study identifies optimal data volumes for accurate diagnosis and offers valuable insights to enhance diagnostic 1 systems. A key focus is on the data samples presented to the classifiers, emphasizing the importance of a reliable sampling strategy to ensure statistically representative samples, particularly for unbalanced datasets, which can significantly improve accuracy and performance. Additionally, univariate and bivariate statistical feature selection techniques further enhance predictive accuracy. This focus on varied sample sizes, coupled with the proposed strategies, presents a novel approach to stroke prediction and broader cardiovascular disease diagnostics.

The paper organises the remaining work: Section II describes related work and explains the complete workflow and the proposed system framework. Section III describes the dataset features and methods. Section IV analyses the results, and describes visualization and interpretation, and Section V summarizes the work and the future scope of the study.

1.1. HEART DISEASES

The world's deadliest disease is heart disease. This disease arises if the heart cannot pump sufficient blood to the different organs [3]. Some heart disease symptoms include weakness, breathing difficulties, and swelling feet. Effective procedures are crucial for identifying complex cardiac disorders that significantly threaten human life [4]. The diagnosis and management of cardiac patients are now complicated due to inadequate physician and diagnostic equipment. Early diagnosis effectively reduces heart-related difficulties and protects against significant dangers [5]. Therapeutic history, specialised symptom analysis, and physical research laboratory testing help surgeons diagnose cardiac abnormalities. Interaction inhibits diagnosis. Age, pulse, and gender might suggest heart illness.

[1,2]*Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi, India*
Emails: 1.vijayalakshmi@bitmesra.ac.in
2.Jayapal@bitmesra.ac.in
[3] *Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi, India*
Email: 3.supreeti.kamilya17@bitmesra.ac.in

Healthcare data analysis helps cardiac patients predict illness, diagnose, assess symptoms, choose drugs, enhance treatment quality, save money, and live longer. Individuals can check for abnormal heartbeat and stroke by placing an ECG sensor on their chest. Advanced clinical data helps physicians predict cardiac disease and make decisions. Human life depends on cardiac blood vessel function. Poor blood circulation can lead to cardiac inactivity, renal failure, and brain imbalance and ultimately result in death. Risk factors comprise BMI, smoking habits, diabetes, high blood pressure, cholesterol, lack of exercise, and inadequate nutrition.

Acute coronary artery spasm is a rare cardiac condition. Arterial spasms appear unexpectedly without atherosclerotic symptoms [6]. Blocking blood flow in the heart leads to low oxygen levels. Women may endure pain for over an hour, whereas men often experience agony for less than an hour. Cardiovascular disease affects the entire body, even distal organs like bone marrow and

spleen [7]. Investigation reveals persistent inflammation. Increased white blood cell counts can lead to inflammation, stroke, and reinfarction [8]. During wound healing, monocytes and macrophages perform inflammatory and reparative functions. Two phases are needed for wound healing; persistent inflammation can induce heart failure.

2. Related Work

Chowdhury et al. [9] presented a machine-learning technique to increase the prediction precision of heart disease by focusing on certain factors. This strategy can highlight certain characteristics. Researchers prepare a questionnaire to predict heart disease. This was accomplished by collecting data from health centres in Bangladesh's Sylhet region. Each of the 564 entries and 18 attributes is included in this dataset. Several machine learning methods, including LR, DT, kNN, SVM, and NB, were evaluated across the dataset. The support vector machine technique outperformed because of its remarkable accuracy rate of 91.0%.

Sangya Ware's study, [10], focuses on diagnosing cardiac illness. The author utilised the Cleveland dataset, which contains information obtained from the UCI Machine Learning repository. These datasets, which contain 303 instances and 14 contributions, are subjected to stringent preprocessing to eliminate noisy data and missing data. LR, NB, kNN, SVM, RF, and DT evaluate six distinct machine-learning algorithms in their research. Model performance was assessed on numerous factors. With an accuracy rating of 89.34%, the Support Vector Machine (SVM) demonstrated exceptional performance.

Li et al. [11] use sophisticated machine-learning methods. In addition to this, traditional methods of attribute selection, such as the highest relevance, the least redundancy, and relief. To solve the issue of feature selection, attribute selection is a method, that is frequently used because it is an extremely effective strategy. SVM has a significantly higher accuracy rate of 92.37% compared to other approaches. This method is highly efficient for identifying heart problems in the healthcare profession.

In [12], various machine learning algorithms, including ANN, SVM, DT, and RIPPER classifiers, are used to predict cardiovascular disease, with, the Cleve land datasets from the UCI library serving as a valuation benchmark. These datasets comprise 303 occurrences and 14 characteristics. Two hundred and ninety-six classifier samples were utilised throughout the preliminary analysis of the data. They explored many classifiers, including ANN, NB, and KNN, and then compared the outcome of these classifiers with the results of the selected technique. The Support Vector Machine (SVM) concluded that the selected approaches performed better, with an accuracy rate of 90.0%.

Lakshmana Rao and colleagues developed a machine learning algorithm to predict cardiovascular disease [13], The evaluation of patients suffering from heart disease is carried out by employing a variety of data mining and neural network techniques. If a diagnosis is delayed, the patient may suffer permanent cardiac damage or perhaps die. Framingham database was utilised to train all of the machine learning algorithms employed. These methods included LR, RF, SVM, DT, KNN, and AdaBoost. Due to its 90.3% accuracy, it is the most effective.

For predicting cardiac disease Hariharan et al. [14] compared machine learning techniques, including KNN, DT, and SVM, The study utilised 270 cases with 12 features from the VA Long Beach dataset, obtained from the University of California, Irvine heart disease repository, to train and evaluate the algorithms. Specificity, sensitivity, and accuracy were assessed, and a confusion matrix was generated based on the test results. The analysis revealed that SVM outperformed DT and KNN in predicting cardiovascular disease (CVD), achieving 83% specificity, 100% sensitivity and 92% accuracy.

A cloud-based system using machine learning was suggested by Nashif et.al [15] to predict cardiac diseases. Researchers integrated two UCI Heart Disease (HD) repository datasets to create a complete dataset. These tactics were implemented using WEKA. The SVM achieved the maximum classification accuracy of 97.53%.

2.1. Research Framework

This research introduces an advanced system for heart disease prediction, leveraging a stacked ensemble approach

to combine and enhance the predictive performance of Support Vector Machine and Logistic regression classifiers. Developed to address the limitations of traditional models, the system is evaluated using the stroke prediction dataset, a benchmark in cardiovascular disease studies. The framework of the system is depicted in Figure 1.

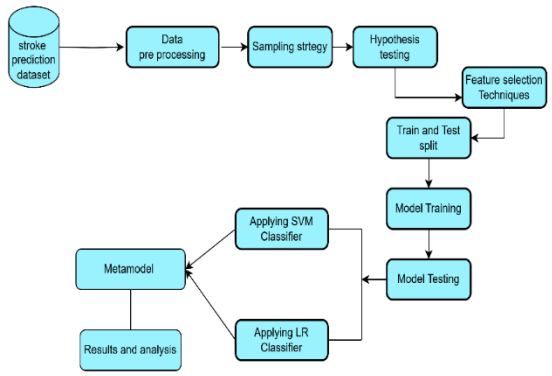

Fig 1: Workflow diagram

3. METHODOLOGY

3.1 Dataset

The study utilized a stroke prediction dataset consisting of 5,110 cases and 12 characteristics, sourced from Kaggle [16]. This dataset comprises 11 features, four demographic and others are clinical attributes, with the target feature, stroke, indicating whether a patient experienced a stroke. Table 1 provides a detailed description of the dataset's characteristics.

3.2 Data preprocessing

A Normalization method of feature scaling is employed to ensure that the data follows a normal distribution. This necessitated using the traditional scaler approach, which involves scaling the output to have unit variance and normalizing each feature by removing its mean. Statistical operations, such as Standard Scalar (SS), have been applied to the dataset to remove the missing values and duplicates from the proposed datasets. The calculation of "unit variance" involves dividing each value by the standard deviation. The procedure is depicted in the following equation, where the new data point (x) is generated by subtracting the mean (μ) value from the previous data point in a designated column and subsequently dividing the outcome by the standard deviation (σ). The standard outcome is as follows.

$$\text{Standard outcome} = \frac{\bar{x} - \mu}{\sigma} \quad (1)$$

3.3 Sampling approach

Two distinct samples of size 397 and 592 are generated from the proposed dataset of size 5110 instances, considering the Z-score at 95% and 99% respectively with a population proportion of 50% and a 5% error margin by using the sample size formula given in equ 1 and then the sample size is adjusted as shown in equation 2

$$n = \frac{Z^2 \cdot p \cdot (1-p)}{ME^2} \quad (2)$$

n = sample size

Z =Z-score corresponding to the confidence level (95% and 99%)

p = population proportion (0.5) ME =margin of error (5%)

$$n_{adj} = \frac{n}{1 + \frac{n-1}{N}} \quad (3)$$

n_{adj} = adjusted sample size ,N =population size (5,110)

3.4 Hypothesis testing

A parametric t-test was performed to determine if a significant difference exists between the sample and population means, based on the null hypothesis that no difference exists. The analysis confirmed that the sample sizes of 357 and 592 observations sufficiently represent the population, as the null hypothesis was upheld. Therefore, these samples are considered representative of the broader population.

3.5 Attribute selection methods

In data analysis, statistical techniques are instrumental in critical attributes that exhibit robust associations with the target variable. The chi-square (chi-square) test is employed initially to select features, concentrating on identifying the most pertinent ones among the non-negative attributes. Following this, Fisher's scores are applied as alternative methods. Collectively, these approaches facilitate data exploration, revealing essential insights and patterns throughout the process.

The significant features selected by the chi-square test and univariate feature selection method are reported in Tables 2 and 3.

Table 1: Description of Variables in the Stroke Prediction Dataset.

Variable	Description
id	Unique identifier for each patient
gender	Categorised as "Male", "Female", or "Other"
age	The patient's age
hypertension	Indicates the presence of hypertension (1) or absence (0)
heart_disease	Indicates the presence (1) or absence (0) of any heart conditions
ever_married	Marital status categorised as "Married"

	or “Not Married”
work type	Characterised as “Children”, “Government Job”, “Never Worked”, “Private Sector”, or “Self-Employed”
Residence type	Categorised as “Urban” or “Rural”
avg glucose level	Mean blood glucose level
bmi	Body Mass Index
smoking status	Smoking status i s categorised as “Formerly Smoked”, “Never Smoked”, “Smokes”, or “Unknown”
stroke	Stroke occurrence indicated by 1 (Yes) or 0 (No)

Table 2: P-values of features using the Chi-Square test

Feature	p-value
heart disease	$2.0677783 \times 10^{-21}$
hypertension	$6.0337512 \times 10^{-23}$
age	$1.9452737 \times 10^{-42}$
bmi	6.4145332×10^{-1}
gender	8.7004085×10^{-1}
work type	1.7055669×10^{-8}
ever married	$3.1283413 \times 10^{-13}$
smoking status	1.2522021×10^{-7}
Residence type	7.2492276×10^{-1}
avg glucose level	1.1796988×10^{-2}

3.6 Statistical Analysis and Interpretations Chi-square Test

Chi-square Test

Chi-square tests for substantial relationships among dataset variables. This investigation studied age, BMI, gender, mode of work, number of married people, type of domicile, average, glucose level, and health problems, including hypertension and heart disease. We aim to find a significant statistical relationship between this test and stroke risk. Demographic and medical data show how many factors affect stroke prediction. These data suggest that age, gender, hypertension, and cardiovascular disease have a significant relationship to stroke occurrence.

Interpretation: Significant features based on the Chi-Square test $p < 0.05$ are heart disease, hypertension, age, work type, ever married, smoking status, and average glucose level.

Fisher Score

The ANOVA-F test is used to select the most vital characteristics. Statistical methods including similarity, dependability, information, and distance reveal substantial input-target correlations in filter-based feature selection. The ANOVA-F test compares each characteristic to the target feature to identify whether they are statistically related. The ANOVA-F test is implemented in Python using the sci-kit-learn’s f classif () method. The SelectKBest function selects the most crucial features (features with the highest scores) using f classif(). The equation to obtain ANOVA-F values is shown in equation 3, and the variance between groups is calculated. The variance within groups is given by equation 4, and the final F-value is computed using equation 5.

$$\text{Variance Between Groups} = \frac{\sum_{i=1}^{S} j_i(\overline{K_i} - \overline{K})^2}{S-1} \quad (4)$$

$$\text{Variance within groups} = \frac{\sum_{i=1}^{S}\sum_{p=1}^{j_i}(\overline{k_{ip}} - \overline{K})^2}{S-1} \quad (5)$$

$$\text{F\_ Value} = \frac{\text{Variance Between Groups}}{\text{Variance within groups}} \quad (6)$$

Table 3: Significant Features ($p \leq 0.05$)

Feature	Score	p-value
age	263.040704	2.620036×10^{-57}
bmi	7.890615	4.993778×10^{-3}
avg glucose level	107.129734	8.664183×10^{-25}
hypertension	47.452936	6.534283×10^{-12}
gender	25.839450	3.883274×10^{-7}
ever married	53.147322	3.723232×10^{-13}
smoking status	14.506882	1.417838×10^{-4}

Interpretation

According to the statistical analysis and interpretation of the above feature selection methods, age and average glucose level significantly affect metabolic and vascular alterations that cause stroke. Through atherosclerosis, atrial fibrillation, and other cardiovascular problems, heart disease discriminates greatly, increasing stroke risk. Smoking, high blood pressure, and BMI all worsen vascular damage, clot formation, and cardiovascular health, raising the risk of stroke. Marriage and gender can influence medical treatment and health behaviors, which raises the risk of stroke.

3.7 Model training

The dataset allocates 80 for training and 20 for testing, ensuring a robust model performance evaluation. We initialize SVM models using the SVC() class without explicitly defining hyperparameters, allowing the models to adapt flexibility to the dataset’s characteristics. We train the

SVM and LR models using the original dataset and its samples. From different statistical perspectives, we consistently identify hypertension, age, work type, ever married, smoking status, average glucose level, and unique features such as BMI, gender, and heart disease as the most significant features. We integrate core features with distinctive features in model training. Since BMI is known to affect disorders such as diabetes and heart disease, it is relevant for projections of health. An individual's pre-existing condition, such as a history of heart disease, could influence their future health risk, and gender shows biological variance in outcomes.

3.8 SVM Classifier

SVM, a supervised machine learning algorithm, is designed for classification tasks, by identifying the optimal hyperplane to separate classes in the feature space. In this context, we use SVM to build a classifier that predicts heart conditions based on the features provided in the dataset. A significant advantage of SVM is its ability to reduce overfitting, particularly in cases with high-dimensional feature spaces. The main aim is to optimise the distance between the hyperplane and the nearest data points from each class. By margin maximisation, SVM seeks to ensure robust and generalised performance across datasets. SVM operates on labelled training data, each data point is represented as a feature vector assigned to one of two classes (binary classification). Each feature vector represents an object with specific features or attributes. In our SVM code, we denoted the features (X) as numerical values that correspond to different characteristics related to heart conditions, the output labels (y) indicate the condition of the heart disease. The algorithm is trained on the labelled data, aiming to maximize the margin between classes while minimizing classification errors.

3.9 LR Classifier

Logistic regression classifiers utilize historical data to estimate binary outcomes, such as 0 or 1. This methodology used a regression model to quantify a dependent variable by examining the relationships between independent variables and creating a binary variable with two possible values 0 or 1. Implementation of Logistic Regression on Proposed Heart Disease Datasets, we established our logistic regression model using the dataset, which includes features such as age, gender, BMI etc. Calculate a linear combination of the input features($x_1,x_2,x_3,x_4,....x_n$) is as follows.

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_n x_n \quad (7)$$

where $\beta_0,\beta_1,\beta_2,...,\beta_n$ are the coefficients (weights) of the features. The sigmoid function $\sigma(z)$ is then applied to the linear combination z to convert it to a probability value between 0 and 1.

$$\sigma(z) = \frac{1}{1+e^{-z}} \quad (8)$$

Thus, the predicted probability $\hat{y}$ can be written as:

$$\hat{y} = \frac{1}{1+e^{-(z)}} \quad (8)$$

Finding the optimal weights that minimize the loss function is vital for training the logistic regression model. The loss function for logistic regression is the log-loss:

$$L(\beta) = -\frac{1}{m}\sum_{i=1}^{n}(y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)) \quad (10)$$

The weights β are optimized using gradient descent. The gradient of the loss function concerning each weight βj is computed as follows:

$$\frac{\partial L(\beta)}{\partial \beta_j} = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)x_{ij} \quad (11)$$

We update the weights using the gradient-descent method as follows:

$$\beta_j \leftarrow \beta_j \leftarrow \alpha\frac{\partial L(\beta)}{\partial \beta_j} \quad (12)$$

Here α is the learning rate.

To make predictions for new patients with feature vector **x**, we calculate

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_n x_n \quad (13)$$

The predicted probability $\hat{y}$ is then:

$$\hat{y} = \frac{1}{1+e^{-(z)}} \quad (14)$$

3.10 Stacked Ensemble Approach:

In the stacked ensemble approach, base models are trained using SVM and LR on the training data. SVM can capture intricate patterns within the data, whereas LR offers a linear approximation of the relationships. Validation data is predicted using trained SVM and LR models. To train a meta-model, the Model Uses the predictions from SVM and LR as input features such as another classifier or regressor. The predictions from these models serve as the input characteristics for the meta-model. It is designed to integrate the predictions generated by the base models and eventually produce the final prediction, marking the culmination of the process.

3.11 Performance Evaluation

To evaluate the performance of SVM, LR, and Stacked ensemble models, F1-score, accuracy, specificity, sensitivity, and precision can be used. These metrics provide significant insights into the model's overall classification performance and suggest information about the accuracy

with which the model differentiates between positive and negative instances.

The following are the performance metrics:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (16)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (17)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (18)$$

F1 score(F1-SCE) represents the harmonic mean of precision and recall. It aims to find an equilibrium between optimizing precision and recall. An F1 score of one signifies the optimal balance, while zero indicates the worst-case scenario (either precision or recall is zero).

$$\text{F1 score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (19)$$

The outcomes of the performance metric assessment are shown in Table 4 and the result is discussed in Section 4.

4 Results and analysis

The analysis of classifier performance on the Stroke prediction dataset for CVD prediction reveals intriguing trends regarding the influence of sample size on predictive efficacy. Initially, The Support Vector Machine (SVM) and LR model were demonstrated on the original dataset and the different sample sizes across various metrics, showcasing, the stronger model is SVM due to its higher sensitivity and better performance metrics in most cases. As sample size rises, LR performs similarly to SVM, but weaker. A stacked ensemble combines SVM and Logistic Regression (LR), Using SVM and LR predictions as input features, a meta-model provides the final prediction. This method uses model strengths to increase prediction performance. The following table 4 shows performance metrics and a graphical representation of performance metrics is referred to in Figure 2.

In a relative analysis of SVM and Logistic Regression models across different dataset sizes, both models exhibit fluctuating stability and adaptability. The SVM shows stable accuracy at 0.76 for a larger sample size, which is an enhancement from 0.74 in the original data, a drop in specificity of 0.65 is observed in sample 1 before improving to 0.74. Sensitivity shows 0.88 in sample size 358 but it regresses to 0.81 in a larger sample, while precision and F1-score drop and then slightly improve, Logistic Regression similarly exhibits a slight decrease in accuracy and specificity in sample size 358, followed by regaining, But there is a consistent decrease in sensitivity across larger sample size. The F1-score declines in both models, showing difficulties in balancing recall and accuracy as sample size increases.

The Stacked Ensemble is applied to control SVM and LR strengths extensively enhancing model performance across various sample sizes and metrics. It attains significantly higher accuracy at 0.96, specificity at 0.95, sensitivity is to 0.97, precision up to 0.95 and F1-score (up to 0.96) compared to SVM and Logistic Regression. Due to the ensemble's capability to combine SVM and LR models' strengths, bias and variance are reduced, and constancy improves across data sizes. The ensemble balanced sensitivity and specificity satisfactorily, addressing the difficulty and unpredictability of higher sample sizes, resulting in an excellent option for applications requiring reliable and accurate predictions.

Table 4: Model Performance Across Different Sample Sizes.

Model	Metric	Original	ample 1	Sample 2
SVM	Acc.	0.74	0.76	0.76
	Spec.	0.73	0.65	0.74
	Sens.	0.81	0.88	0.81
	Prec.	0.94	0.78	0.81
	F1-scr.	0.81	0.76	0.78
LR	Acc.	0.75	0.73	0.76
	Spec.	0.74	0.67	0.75
	Sens.	0.81	0.78	0.78
	Prec.	0.94	0.78	0.81
	F1-scr.	0.82	0.73	0.74
Stacked Ensemble	Acc.	0.90	0.93	0.96
	Spec.	0.89	0.89	0.95
	Sens.	0.91	0.97	0.97
	Prec.	0.89	0.89	0.95
	F1-scr.	0.93	0.93	0.96

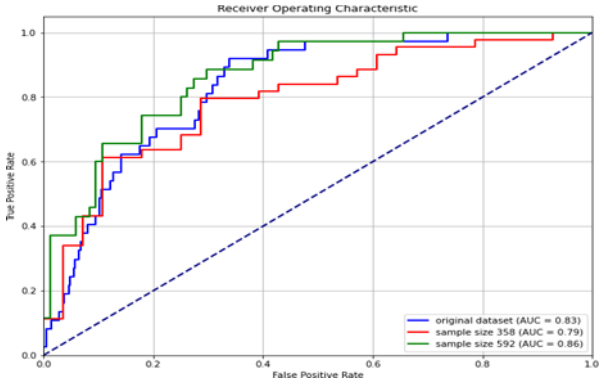

Fig 2: Performance of SVM, LR and Ensemble classifier.

C-statistic

C-statistics increase, and the model's ability to distinguish between positive and negative classes improves. This statistic is particularly beneficial for datasets with a high-class imbalance, providing vital insights into performance evaluation. A 0.5 score indicates random performance but values 0.6 to 0.7 show poor performance, values between 0.7 to 0.8 show fair discrimination, and values between 0.8 to 0.9 are good performance. The "area under the curve" (AUC) for each observation is the percentage likelihood that a classifier would score a positive observation higher than a negative one.

C-statistic is a significant parameter for testing classification models, notably in medical fields like stroke prediction. The ensemble classifier for the stroke prediction dataset has AUC values of 0.90 for the original dataset, 0.93 for 358 samples, and 0.96 for 592 samples. These data show that the model can accurately identify stroke-prone people. A moderately dependable model with an AUC of 0.90 indicates a 90% accuracy, in categorizing people. This dependability rises with sample number, reaching 0.93 with 358 samples and 0.96 with 592 samples, demonstrating model performance improvement with additional data. Improved AUC indicates a model's greater accuracy and reliability in forecasting critical situations, boosting its application in healthcare for early intervention and treatment planning. In medical prediction tasks, rigorous data collection and model modification are crucial to optimum performance. Larger datasets increase AUC values. and graphical representation is shown in Figures 3 and 4.

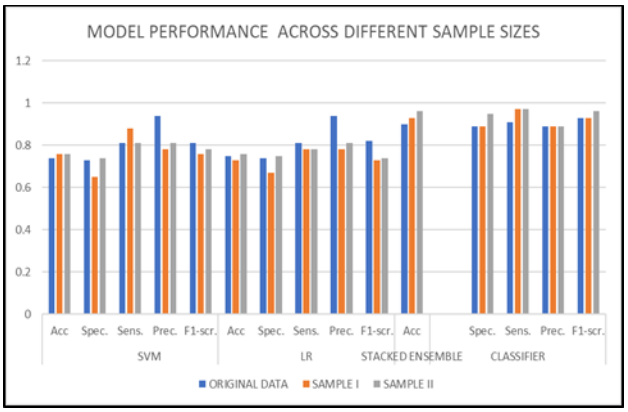

Fig 3: SVM and LR OF AUC and ROC

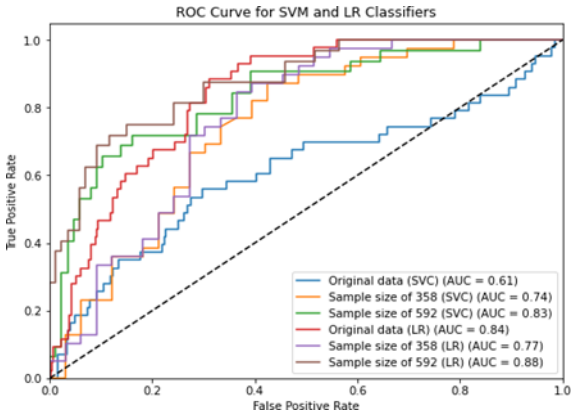

Fig 4: AUC of Stacked Ensemble classifier

5. Conclusion

SVM, Logistic Regression, and Stacked Ensemble approaches on varied medical datasets provide healthcare predictive modelling insights. The Stroke Prediction dataset's SVM has better sensitivity with larger sample sizes, but greater specificity with smaller samples, indicating a higher false positive rate. The heart disease stroke prediction dataset's Logistic Regression performs consistently across sample sizes, indicating dependability. Our Stacked Ensemble model exceeds the original stroke prediction dataset with higher accuracy (96% for sample 2, 93% for sample 1) and comparable sensitivity, specificity, and F1-score when applied to larger samples.

Future directions for this research include enhancing the robustness of models by incorporating more base models on varied datasets and delving into advanced feature engineering to discover latent predictive patterns. These advancements hold the potential to significantly impact medical diagnostics, leading to improved patient care and more efficient use of healthcare.

References

[1] S. Jain and A. Saha, "Rank-based univariate feature selection methods on machine learning classifiers for

code smell detection," Evolutionary Intelligence, Vol. 15, no. 1, 2022, pp. 609–638.

[2] Y. Liu, Y. Zhou, S. Wen, and C. Tang, "A strategy on selecting performance metrics for classifier evaluation," International Journal of Mobile Computing and Multimedia Communications (IJMCMC), Vol. 6, no. 4, 2014, pp.20–35.

[3] A. L. Bui, T. B. Horwich, and G. C. Fonarow, "Epidemiology and risk profile of heart failure," Nature Reviews Cardiology, Vol. 8, no. 1, 2011, pp.30–41.

[4] J. Mourao-Miranda, A. L. Bokde, C. Born, H. Hampel, and M. Stetter, "Classifying brain states and determining the discriminating activation patterns: support vector machine on functional mri data," Neuroimage, Vol. 28, no. 4, 2005, pp. 980–995.

[5] S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, "Innovative artificial neural networks-based decision support system for heart diseases diagnosis," 2013.

[6] J. Soni, U. Ansari, D. Sharma, S. Soni et al., "Predictive data mining for medical diagnosis: An overview of heart disease prediction," International Journal of Computer Applications, Vol. 17, no. 8, 2011, pp. 43–48.

[7] H. D. Masethe and M. A. Masethe, "Prediction of heart disease using classification algorithms," in Proceedings of the World Congress on Engineering and Computer Science, Vol. 2, no. 1, 2014, pp. 25–29.

[8] K. Patterson, "Matthiasnahrendorf: Healing insights," Circulation Research, Vol. 119, no. 7, 2016, pp. 790–793. 15

[9] M. N. R. Chowdhury, E. Ahmed, M. A. D. Siddiq, and A. U. Zaman, "Heart disease prognosis using machine learning classification techniques," in 2021 6th International Conference for Convergence in Technology (I2CT). IEEE, 2021, pp. 1–6.

[10] S. Ware, S. Rakesh, and B. Choudhary, "Heart attack prediction by using machine learning techniques," no, Vol. 5, 2020, pp. 1577–1580.

[11] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart disease identification method using machine learning classification in healthcare," IEEE Access, Vol. 8, 2020, pp. 107 562–107 582.

[12] S. N. Khan, N. M. Nawi, A. Shahzad, A. Ullah, M. F. Mushtaq, J. Mir, and M. Aamir, "Comparative analysis for heart disease prediction," JOIV: International Journal on Informatics Visualization, Vol. 1, no. 4-2, 2017,pp. 227–231.

[13] A. Lakshmanarao, Y. Swathi, and P. S. S. Sundareswar, "Machine learning techniques for heart disease prediction," Forest, Vol. 95, no. 99, 2019, p. 97.

[14] K. Hariharan, W. Vigneshwar, N. Sivaramakrishnan, and V. Subramaniyaswamy, "A comparative study on heart disease analysis using classification techniques," International Journal of Pure and Applied Mathematics, Vol. 119, no. 12, 2018, pp. 13 357–13 366.

[15] S. Nashif, M. R. Raihan, M. R. Islam, and M. H. Imam, "Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system," World Journal of Engineering and Technology, Vol. 6, no. 4, 2018, pp. 854–873.

[16] Fede Soriano, "Stroke prediction dataset," Kaggle,2023,https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset Accessed: 2024-06-18.