JOURNAL MEDICAL INFORMATICS TECHNOLOGY
medinftech.org

Logistic Regression with Hyper Parameter Tuning Optimization for Heart Failure Prediction

Teguh Herwanto[1], Wan Ahmad Gazali Kodri[2], Faruq Aziz[3], Alya Shafira Hewiz[4], and Dwiza Riana[5]*

[1,2,3,5] *Universitas Nusa Mandiri Jakarta, Indonesia*
[4] *Universitas Airlangga, Surabaya*

ARTICLE HISTORY

Received: 13 March 23

Final Revision: 16 March 23

Accepted: 30 March 23

Online Publication: 31 March 23

KEYWORDS

Logistic Regression, Hyper Parameter Tuning, Feature Selection, Heart Failure Prediction, RapidMiner

CORRESPONDING AUTHOR

14220001@nusamandiri.ac.id

DOI

10.37034/medinftech.v1i1.3

A B S T R A C T

Heart failure is a major public health concern that causes a substantial number of deaths worldwide. Risk factor analysis is required to diagnose and treat patients with heart failure. The logistic regression with hyper parameter tuning optimization is presented in this research, with ejection fraction, high blood pressure, age, and serum creatinine as relevant risk factors. This study indicates that better data preparation utilizing Deep Learning with hyper parameter adjustment be used to determine the best parameter that has a substantial influence as a risk factor for heart failure. The experiments employed data from the Faisalabad Institute of Cardiology and Allied Hospital in Faisalabad (Punjab, Pakistan), which included 299 samples. The experimental results show that the proposed approach outperforms related studies with a recall of 63.16%.

1. Introduction

Heart disease is a critical public health issue that affects people all over the world. According to the World Health Organization (WHO), cardiovascular disease (CVD), which includes coronary heart disease, stroke, and heart failure, is responsible for a substantial number of deaths worldwide [1]. The World Health Organization reports that CVD, including heart failure, accounts for 31% of all global mortality [2]. Heart failure occurs when the heart is unable to properly pump blood to meet the body's demands. High blood pressure, diabetes, coronary heart disease, and other heart problems can all increase the risk for getting heart failure. CVD, particularly heart failure, is one of the leading causes of death worldwide [3].

To maintain optimal function of existing tissues, the human body requires an adequate mix of oxygen and nutrients. The heart is the primary pump that circulates blood throughout the body via the circulatory system. If the heart's function is disrupted, the circulatory system can become clogged, resulting in heart failure [4].

Ejection fraction, high blood pressure, serum creatinine, anemia, diabetes, creatinine phosphokinase, platelets, serum sodium and other parameters are involved. These indicators have a strong correlation in heart failure patients.

Furthermore, age-related factors influence the susceptibility to heart failure. Identification and extraction of some vital information about these factors could lead to better heart disease prediction [5], which is an important component of efforts to prevent, manage, and treat heart disease. Machine learning can be used to assess and identify people with heart failure, and it is widely employed in research and prediction. [6]. As a result, it encourages researchers to identify features that have a significant impact on the prediction of heart failure.

This study proposes a model for detecting heart failure that takes into consideration ejection fraction, high blood pressure, serum creatinine, anemia, creatinine phosphokinase, platelets, serum sodium, age, and diabetes. The study also looks into the optimum approach for imbalance data by using existing sampling methods such as the Synthetic Minority Over Sampling Technique (SMOTE). By incorporating known feature selection methods such as Forward Selection, this experiment aims to determine the best parameter that has a significant impact on the best prediction. This study is predicted to produce the best predictive model

with high accuracy and sensitivity when compared to previous work.

2. Research Method

2.1. Related Work

The same dataset is used in [7], and they apply several machine learning classifiers to both predict the patient's survival and rank the features corresponding to the most important risk factors. They only used serum creatinine and ejection fraction to predict mortality events using feature ranking analysis and according to results from traditional biostatistical tests. The logistics regression method is the best model to predict heart failure with accuracy 58.5%, and sensitivity 54.1%.

Using the same dataset [8], the random forest method was used to predict death occurrences and implementing the existing sampling methods such as Synthetic Minority Over Sampling Technique (SMOTE), Borderline-SMOTE, and adaptive synthetic (ADASYN) to balance unbalanced classes in the dataset. The experimental results show that ejection fraction, serum creatinine, and age are all highly significant factors for predicting heart failure. Their findings yielded a sensitivity of 71.23%, a specificity of 75.11%.

Table 1. Faisalabad Heart Failure Dataset

No	Attribute	Description
1	Age	Age of the patient (years)
2	Anemia	Decrease of red blood cells or Hemoglobin (boolean)
3	High Blood Pressure	If the patient has hypertension (boolean)
4	Creatinine Phosphokinase (CPK)	Level of the CPK enzyme in the blood (mcg/L)
5	Diabetes	If the patient has diabetes (boolean)
6	Ejection Fraction	percentage of blood leaving the heart at each contraction (percentage)
7	Platelets	platelets in the blood (kiloplatelets/mL)
8	Serum Creatinine	level of serum creatinine in the blood (mg/dL)
9	Serum Sodium	level of serum sodium in the blood (mEq/L)
10	Sex	woman or man (binary)
11	Smoking	yes or no (binary)
12	Time	follow-up period (days)
13	Death Event	if the patient deceased during the follow-up period (boolean) as a target attribute

Other research was conducted by using two primary features: serum creatinine and ejection fraction in early stage [9] from the same dataset. After that, two feature selection approaches, minimum redundancy maximum relevance, and recursive feature elimination based on Naïve Bayes were employed. Then employs a decision tree technique for the best classifiers. They achieved an accuracy of 80%, a sensitivity of 51.72%, and a specificity of 93.44%.

Previously, numerous machine learning algorithms or feature selection approaches [10] and handle unbalanced class were used in the study. There are quite a few research have employed a combination of feature selection, SMOTE and hyper parameter optimization approaches before. This study proposes SMOTE to find out the best sampling class. Then, a pre-processing feature selection based on Deep Learning is utilized to pick the best and most significant features. On the other hand, hyper parameter tweaking is used to optimize the prediction model generated by the classification process using logistic regression.

2.2. Proposed Method

All the experiments have been conducted in a Rapid Miner Studio v.10.1 Educational Edition on Intel Core i5 2 Cores running at 2.4Ghz machine which is equipped with 16 GB DDR4 Random Access Memory (RAM). Figure 1 depicts the overall research technique.

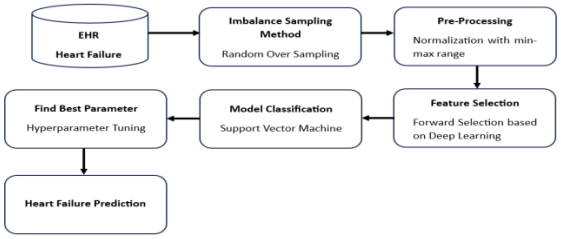

Figure 1. Block diagram of heart failure prediction

2.3. Dataset

This dataset contains medical information from 299 Electronic Health Record (EHR) patients who were treated at the Faisalabad Institute of Cardiology and Allied Hospital in Faisalabad (Punjab, Pakistan) for heart failure. These patients ranged in age from 40 to 95 years and included 105 women and 194 males. According to the New York Heart Association (NYHA) categorization for stages of heart failure, all 299 patients had ventricular systolic dysfunction and previous heart failure, putting them in class III or IV. During the follow-up period, 96 patients died as a result of heart failure. As stated in Table 1, this dataset comprises 13 attributes, one of which is a target attribute.

In general, raw data frequently contains noisy and inconsistent examples, which might impede effective prediction and proper machine learning analysis [11]. As a result, data preparation is a key stage in machine learning analysis since it prepares the data set for

improved analysis outcomes. The heart failure dataset is preprocessed in this context. Additionally, feature engineering is performed to change the type of feature if necessary.

We removed the sex and time attributes since they could lead to biased models or decisions. The death event attribute has a value of 1 for patients who died and 0 for those who survived the treatment term. We defined a positive class (class 1) based on the death event attribute of the deceased patients (death event = 1). We divided the dataset into two sections, 80% of the total amount for the training set and 20% for the test set. To ensure that the results are reproducible, we select 1992 as the local random seed.

2.4. Imbalanced Sampling Method

Synthetic Minority Over-Sampling Technique (SMOTE) is a random over sampling method that generates data on the kind portion connecting minority class sample and its k-Nearest Neighbors (KNN) [8]. First, random data is selected from a minority class sample. Next, the KNN are randomly chosen. In SMOTE, a new synthetic minority class x_{new} is generated, which lies on the line segment between x_i and x_k as Equation (1).

$$x_{new} = (x_i - x_k) \times \delta \quad (1)$$

Where x_i is minority class random data, k is hyper-parameter of KNN, x_k is KNN of x_i and δ is random value between 0 and 1.

We are randomly selecting examples from the training set, and the sample size is based on a relative random sample with class 1 by ratio 1.0 and class 0 by ratio 0.9. Then we apply SMOTE to these random examples and set the number of neighbors to two.

2.5. Pre-Processing

We increased machine learning accuracy using min-max data standardization to compensate for the dataset's small size, and the model may then be trained using the dataset [12]. It is important to standardize the data before forecasting with machine learning. Normalization is required because some machine learning algorithms want to detect patterns in data sets by comparing data point properties. If the qualities are on significantly different scales, there will be issues. Simply expressed, when numerous properties have values at different sizes, a weak data model can arise. Min-max normalization is the normalizing approach we use to ensure that all attributes or features have the same scale [13].

Min-max normalization is a linear data normalization procedure in which the minimum data value becomes 0 and the maximum data value becomes 1 for each variable. Equation (2) below is used to modify each value [14].

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2)$$

Where x' is new value of each entry, x is attribute data values, $\max(x)$ is maximum value of x, and $\min(x)$ is minimum value of x.

Bias can occur when variables or attributes considered at different levels do not contribute equally to the fit of the model and function being trained. As a result, feature-based normalization, such as min-max scaling, is frequently used prior to model fitting to mitigate this potential issue. As a result, the data is not cleansed or sorted prior to normalization. We have data that is neatly aligned and well-ordered or normalized after min-max normalization. The accuracy of machine learning models improves when data is always standardized.

2.6. Feature Selection

In general, a dataset comprises numerous sorts of valuable information that cannot be easily retrieved. This information can be derived by studying the dataset. We employed statistics and Exploratory Data Analysis (EDA) on the dataset in this work to reveal hidden patterns and trends [15]. It helps to qualitatively get more information about the dataset by visualizing the attributes characteristics [16] or primarily through the use of graphics. Also, EDA was used to assess the data prior to modeling [17].

Table 2. Parameter of Feature Selection

Selected Parameters	Value
Selection Direction	forward
Limit Generations Without Improval	1
Maximum Number Of Generations	30
Keep Best	3

A Deep Learning approach based on forward selection was employed with a tuning hyperparameter to apply feature selection. Table 2 shows how we use the feature selection parameter, and Table 3 shows how we tune the hyper parameters. As demonstrated in Table 4, Deep Learning clearly identifies ejection fraction, high blood pressure, age, and serum creatinine as the most important and meaningful indicators in predicting mortality or survival in heart failure patients.

Table 4 depicts the four most significant features that are responsible for heart failure, survival, and mortality. This is the most important risk factor for heart failure patients. Ejection fraction has significantly positive for heart failure prediction, while high blood pressure, age and serum creatinine are negatively affects for heart failure prediction.

Table 3. Hyper Parameters of Deep Learning

Selected Parameters	Value
Nesterov_Accelerated_Gradient	True
Momentum_Start	0.9
Learning_Rate	0.01
Activation	Tanh
Learning_Rate_Decay	0.5

Table 4. Result of Feature Selection

Attribute	Weight
Ejection_Fraction	1.026
High_Blood_Pressure	-0.399
Age	-0.585
Serum_Creatine	-1.259

2.7. Logistic Regression

The relationship between one binary dependent variable, also known as the outcome variable, and one or more independent variables, also known as the covariate or explanatory variable, is described and estimated using logistic regression models. Logistic regression models are adaptable, have powerful interpretations, and have been utilized to analyze events in a variety of medical and non-medical research disciplines. Logistic regression models, like other types of regression models, are frequently used to evaluate predictors and account for variables and/or relationships.

These models are used to examine retrospective data, including case-control studies, and to construct predictive algorithms that can be presented in nomograms or online calculators to express the probability of an occurrence, such as toxicity or light bulb failure in the two instances above. When comparing two cohorts on an outcome, logistic regression models are also used to generate propensity scores, which can subsequently be used to balance the case mix (i.e., a variety of characteristics) of the two cohorts. In unpaired or weighted analysis, likelihood ratings for logistic regression models are commonly utilized. Finally, logistic regression is a common machine learning approach for two-class classification problems since the outcome variable is binary [18].

In this study, we used the default values for Logistic Regression parameters and subsequently improved them via hyper parameter optimization.

2.8. Hyper Parameter Optimization

Optimized Parameters (Grid) was the optimization approach used in this study. The goal is to find the best model performance combination that can be used as an logistic regression prediction model.

Table 5. Logistic Regression Hyper Parameters

Selected Parameters	Value
Solver	IRLSM
Reproducible	true
Maximum_Number_Of_Threads	min=1, max=100, steps=10
Compute_P-Values	true
Add_Intercept	true
Max_Iterations	min=0, max=100, steps=10
Remove_Collinear_Columns	true
Regularization	true
Lambda	min=0.1, max=100, steps=2
Lambda_Search	true
Alpha	min=0, max=1, steps=2
Early_Stopping	true
Stopping_Rounds	3
Stopping_Tolerance	0.001

Table 5 illustrates how we use hyper parameter tuning. Other research results show that hyper parameter tuning can improve accuracy, recall, precision, and AUC [19].

2.9. Performance Evaluation Criteria

Accuracy, recall/sensitivity, specificity, G-mean, Matthews's correlation coefficient (MCC), and receiver operating characteristic area under the curve (AUROC) were used in this study to determine the classification algorithm with the best performance. The coefficient value is used to evaluate risk variables. It highlights key risk factors. The following equation is used to calculate the value of this measure [20], [21], [22], [23], [24].

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (3)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (4)$$

$$Specificity = \frac{TN}{TN+FP} \quad (5)$$

$$G-mean = \sqrt{Sensitivity \times Specificity} \quad (6)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (7)$$

True positives, false positives, false negatives, and true negatives are represented by the letters TP, FP, FN, and TN. The ROC curve is created using the value received from the confusion matrix computation, which is the difference between the False Positive Rate (1 - Specificity) and the True Positive Rate (Sensitivity) [25].

$$True\ Positive\ Rate = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$False\ Positive\ Rate = \frac{False\ Positive}{False\ Positive + True\ Negative}$$

Sensitivity or Recall, also known as True Positive Rate (TPR), represents the performance of the classifier on the positive (minority) class. A higher value of sensitivity reflects that the classifier is good at predicting the minority class instances.

We want to avoid false negatives as much as possible when predicting heart failure. A false negative case suggests that a heart disease problem was missed, which could be fatal. False positives (a healthy person with a false heart problem diagnosis) are not as critical in this use case as preventing a circumstance in which we do not detect a major problem such as heart disease.

If a person has heart disease, he or she should not be considered a healthy patient. Undiagnosed heart disease should not result in an accident that stops the patient from receiving proper care. In other words, recall becomes a critical metric.

3. Result and Discussion

The datasets summarize the results of the descriptive statistical analysis of the 299 heart failure dataset

observations, 96 of which were heart failure deaths and 203 of which were heart failure survivors.

Table 6. Selection of Comparison for Logistic Regression

Evaluation	Without SMOTE and Feature Selection	SMOTE	SMOTE and Feature Selection
Accuracy	68.33	31.67	**76.67**
G-Mean	0	0	70.30
Sensitivity	0	100	57.89
Specificity	100	0	85.37
AUC	50	0	75
MCC	-	-	44.65

The model was tested without using the sample class imbalance approaches technique, and feature selection was performed initially for comparison with SMOTE methods and feature selection. Table 6 lists the experimental outcomes in terms of accuracy, G-Mean, sensitivity, specificity, AUC, and MCC, and Figure 2 shows a visual comparison of them. The best accuracy for SMOTE and Feature Selection was reached by applying logistic regression, which had an average accuracy of 76.67%.

The importance of individual features for each classifier is established via forward elimination, depending on the value of the importance of the features found by each applicable machine learning method. Feature significance values were determined to find the features that are most essential in predicting the chance of death in heart failure.

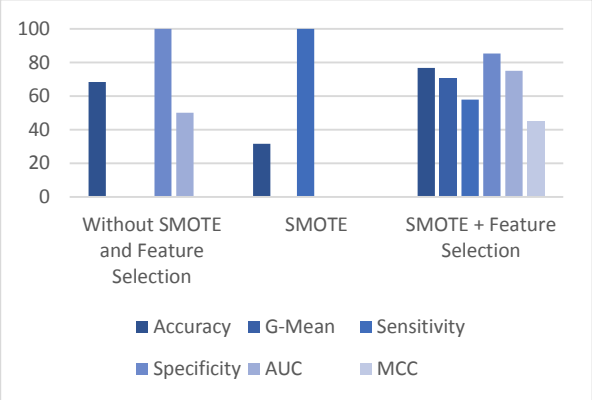

Figure 2. Examining Evaluation Metrics

Table 7. Comparison Logistic Regression with Optimization

Evaluation	SMOTE and Feature Selection	SMOTE and Feature Selection with Hyper Parameters
Accuracy	76.67	73.33
G-Mean	70.30	70.21
Sensitivity	57.89	**63.16**
Specificity	85.37	78.05
AUC	75.00	72.80
MCC	44.65	40.19

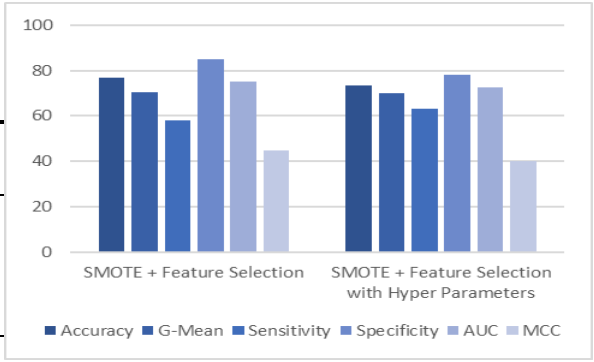

Figure 3. Comparison of Logistic Regression with Optimization

Our proposed logistic regression classifier was trained using only SMOTE, feature selection. When the approach's current logistic regression hyperparameter was tuned, the classifier's prediction performance improved noticeably. The results of incorporating hyper parameter optimization methodologies are shown in Table 7 and Figure 3.

When compared to earlier relevant studies, Table 8 shows that the suggested technique, logistic regression (LR) using SMOTE, feature selection (FS), and hyper parameters optimization (HYPER), yields the best results with an AUC of 72.80% and a sensitivity of 63.16%. Some measures were not reported by the authors. Therefore, they are kept empty, such as AUC and MCC metric. Figure 4 depicts a visual comparison with related research studies.

Table 8. Performance Evaluation Against Related Research (%)

Study	Chicco [7]	Kim [8]	Hasan [9]	This Study
Method	RF univariate feature selection	RF +SMOTE	DT +MRMR +RFE	LR +SMOTE +FS +HYPER
Accuracy	58.50	–	80.00	73.33
G-Mean	68.01	73.14	69.52	70.21
Sensitivity	54.10	71.23	51.72	**63.16**
Specificity	85.50	75.11	93.44	78.05
AUC	69.80	–	72.58	**72.80**
MCC	41.80	–	–	40.19

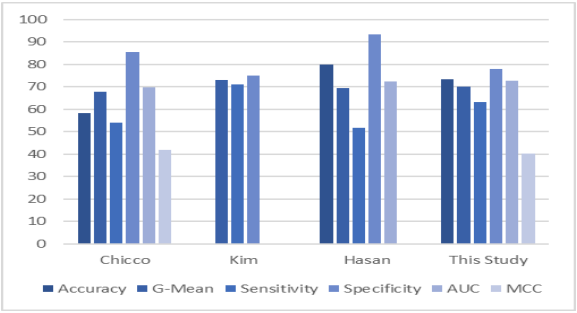

Figure 4. Performance Evaluation Against Related Research

4. Conclusion

The study's goal is to create a machine learning model that can predict heart failure survival and identify the most important risk variables. Overall, the logistic regression with SMOTE hyper parameter tweaking and

feature selection outperforms previous models such as DT + MRMR + RFE. Ejection fraction, high blood pressure, age, and serum creatinine have the potential to become key markers for predicting heart failure. Our proposed model could be useful for physicians and experts screening heart failure patients in a clinical setting. SMOTE hyper parameter tweaking, and Deep Learning feature selection achieved an accuracy of 73.33%, AUC of 72.80%, and, most importantly in this case, recall or sensitivity of 63.16%. The importance of recall or sensitivity metric is to ensure that we will identify all of the heart disease patients quickly and could be determinant by four attribute form feature selection result so that the patients will receive correct care in follow-up clinical examination. In the future, we plan to collect larger and more diverse data sets and build using cutting-edge technologies such as transfer learning.

References

[1] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," BMC Med Inform Decis Mak, vol. 20, no. 1, Feb. 2020, doi: 10.1186/s12911-020-1023-5.

[2] World Health Organization, "Cardiovascular diseases (CVDs)," *https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)*.

[3] L. and B. I. National Heart, "What Is Heart Failure, 2022," *https://www.nhlbi.nih.gov/health/heart-failure*.

[4] Uci.edu, "UCI Machine Learning Repository." https://archive.ics.uci.edu/ml/datas ets/Heart+failure+clinical+records (accessed May 11, 2023).

[5] T. Milo and A. Somech, "Automating Exploratory Data Analysis via Machine Learning: An Overview," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Association for Computing Machinery, Jun. 2020, pp. 2617–2622. doi: 10.1145/3318464.3383126.

[6] G. Chandrashekar and F. Sahin, "32 A survey on feature selection methods," *Computers and Electrical Engineering*, vol. 40, no. 1, pp. 16–28, Jan. 2014, doi: 10.1016/j.compeleceng.2013.11.024.

[7] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC Med Inform Decis Mak*, vol. 20, no. 1, Feb. 2020, doi: 10.1186/s12911-020-1023-5.

[8] Kim YT, Kim DK, Kim H, and Kim DJ, 29 A comparison of oversampling methods for constructing a prognostic model in the patient with heart failure. 2020. doi: https://doi.org/10.1109/ICTC49870.2020.9289522.

[9] M. Al Mehedi Hasan, J. Shin, U. Das, and A. Yakin Srizon, "Identifying Prognostic Features for Predicting Heart Failure by Using Machine Learning Algorithm," in ACM International Conference Proceeding Series, Association for Computing Machinery, Mar. 2021, pp. 40–46. doi: 10.1145/3460238.3460245.

[10] A. Newaz, N. Ahmed, and F. Shahriyar Haq, "Survival prediction of heart failure patients using machine learning techniques," *Inform Med Unlocked*, vol. 26, Jan. 2021, doi: 10.1016/j.imu.2021.100772.

[11] P. Ghosh *et al.*, "Efficient prediction of cardiovascular disease using machine learning algorithms with relief and lasso feature selection techniques," *IEEE Access*, vol. 9, pp. 19304–19326, 2021, doi: 10.1109/ACCESS.2021.3053759.

[12] A. C. M. da Silveira, Á. Sobrinho, L. D. da Silva, E. de B. Costa, M. E. Pinheiro, and A. Perkusich, "Exploring Early Prediction of Chronic Kidney Disease Using Machine Learning Algorithms for Small and Imbalanced Datasets," *Applied Sciences (Switzerland)*, vol. 12, no. 7, Apr. 2022, doi: 10.3390/app12073673.

[13] L. Munkhdalai, T. Munkhdalai, K. H. Park, H. G. Lee, M. Li, and K. H. Ryu, "Mixture of Activation Functions with Extended Min-Max Normalization for Forex Market Prediction," *IEEE Access*, vol. 7, pp. 183680–183691, 2019, doi: 10.1109/ACCESS.2019.2959789.

[14] Vatsal Gajera, Shubham, Rishabh Gupta, and Prasanta K. Jana, An Effective Multi-Objective Task Scheduling Algorithm using Min-Max Normalization in Cloud Computing.

[15] W.J. Dixon and F.J. Massey Jr, "Introduction to statistical analysis," 1951.

[16] A. Ishaq et al., "Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques," in IEEE Access, vol. 9, pp. 39707-39716, 2021, doi: 10.1109/ACCESS.2021.3064084.

[17] J. W. Tukey, "Exploratory Data Analysis," vol. 2, pp. 131–160, 1977.

[18] E. C. Zabor, C. A. Reddy, R. D. Tendulkar, and S. Patil, "Logistic Regression in Clinical Studies," *Int J Radiat Oncol Biol Phys*, vol. 112, no. 2, pp. 271–277, Feb. 2022, doi: 10.1016/j.ijrobp.2021.08.007.

[19] Ernawan, Ferda & Handayani, Kartika & Fakhreldin, Mohammad & Abbker, Yagoub. (2022). Light Gradient Boosting with Hyper Parameter Tuning Optimization for COVID-19 Prediction. International Journal of Advanced Computer Science and Applications. 13. 514-523. 10.14569/IJACSA.2022.0130859.

[20] Md Mamun Ali *et al.*, "Machine learning-based statistical analysis for early stage detection of cervical cancer," *Comput Biol Med*, vol. 139, 2021.

[21] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn, and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison," *Comput Biol Med*, vol. 136, Sep. 2021, doi: 10.1016/j.compbiomed.2021.104672.

[22] Asifa Afsari Hemu, Rabeya Begum Mim, Md Mamun Ali, Md. Nayer, Kawsar Ahmed, and Francis M. Bui, "Identification of Significant Risk Factors and Impact for ASD Prediction among Children Using Machine Learning Approach," in *Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, 2022.

[23] Md. Sazzad Hossain, Mehedi Hassan Shovo, Md. Mamun Ali, Md. Nayer, Kawsar Ahmed, and Francis M. Bui, "ML_SPS: Stroke Prediction System Employing Machine Learning Approach," in *Artificial Intelligence and Data Science*, 2022, pp. 215–226.

[24] K. Ahmed and M. Z. Rahman, "Early Prevention and Detection of Skin Cancer Risk using Data Mining," 2013.

[25] R. Vankdothu and M. A. Hameed, "Brain tumor MRI images identification and classification based on the recurrent convolutional neural network," *Measurement: Sensors*, vol. 24, Dec. 2022, doi: 10.1016/j.measen.2022.100412.