Gradient Boosting Algorithm with Logistic Regression Estimation of Accuracy Rate in Predicting Cardiovascular Disease.

Talluri Vishnuvardhan[1], A. Rama[2*]

[1]Research Scholar, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, 602105.

[2*]Project Guide, Corresponding Author, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, 602105.

ABSTRACT:

Aim: Estimation of accuracy rate in predicting the cardiovascular disease using gradient boosting algorithm with novel logistic regression. **Materials and Methods:** The prediction is done by Novel ANN (N=10) and KNN (N=10) algorithms. The sample size is determined using the G power Calculator and it's found to be 10. **Results:** Based on the results accuracy obtained in terms of accuracy is identified by Gradient boosting (91.00%) over the Logistic Regression algorithm (92.18%). The statistical significance difference between the novel gradient boosting algorithm and Logistic Regression Algorithm was found to be p=0.001 (2 tailed) (p<0.05). **Conclusion:** Prediction of cardiovascular disease using Logistic Regression is significantly better than the Gradient Boosting Algorithm.

Keywords: Novel gradient boosting, Novel logistic regression, Machine learning, Diet, Cancer, Obesity.

INTRODUCTION

Cardiovascular sickness has been viewed as the most serious, furthermore deadly sickness in people. The expanded pace of cardiovascular illnesses with a high death rate is causing critical danger and weight to the medical healthcare frameworks around the world. Cardiovascular illnesses are more found in men than in ladies especially in the center or advanced age in spite of the fact that there are additional youngsters with comparative medical healthcare problems. According to the information given by the WHO, 33% of the deaths internationally are brought about by coronary illness. CVDs cause the passing of roughly 17.9 million individuals consistently worldwide and have a higher predominance in Asia. The European Cardiology Society ESC detailed that 26 million grown-ups worldwide have been determined to have coronary illness, and 3.6 million are recognized every year. Generally, 50% of all patients determined to have Heart Disease bite the dust within 1-2 years and around 3% of the complete financial plan for medical healthcare services is conveyed on treating coronary illness Obesity (Yadav and Pal 2021). To foresee coronary illness different tests are required. The absence of skill of clinical staff may bring about misleading forecasts. Early conclusions can be troublesome diet (Sahay et al. 2022). Careful treatment of coronary illness is testing, especially in

non-industrial nations which need prepared clinical staff as well as testing hardware and other assets expected for legitimate determination and care of patients with heart issues (Mendis et al. 2011). A precise assessment of the danger of cardiovascular disappointment would assist with forestalling serious respiratory failures and furthermore work on the well-being of patients DietinObesity (Nesaragi et al. 2022)). Calculations can be viable in distinguishing the infections when prepared on legitimate information Diet in Healthcare. The Real World Applications of Logistic Regression are Medicine, Text Editing, Hotel Booking, etc. Coronary illness datasets are openly accessible for the examination of forecast models (Govoni, Politi, and Vanoli 2020).

Presentation of AI and computerized reasoning assists the specialists with planning the best forecast model utilizing the enormous data sets which are accessible to Cancer (Mehta and McSweeney 2018). Late examinations which center around the heart-related issues in grown-ups and kids underscored the need of decreasing mortality connected with CVDs. Since the accessible clinical datasets are conflicting Cancer (Trevisan, Sergi, and Maggi 2020). What's more, appropriate preprocessing is a vital advance (Tavares et al. 2021). Choosing the critical elements that can be utilized as the hazard factors in expectation models is fundamental Obesity (B. and Manikandan 2021). Care ought to be taken to choose the right mix of highlights and the proper AI calculations to create precise expectation models Diet (Mackay et al. 2004). It is critical to assess the impact of hazard factors that meet the three rules like the high pervasiveness in many populaces (Sahay et al. 2022); a critical effect on heart sicknesses freely; and they can be controlled or treated to lessen the dangers is Obesity (Mendis et al. 2011). Various specialists have included different danger factors or elements while displaying the indicators for CVD Cancer (Wiharto, Wiharto, Herianto, et al. 2017). In the Last 5 years, 2017-2021 Google Scholar has published more than 196 papers and the IEEE published more than 200 papers about the Prediction of cardiovascular disease. The analysis of the Novel Gradient boosting Algorithm and novel Logistic Regression Algorithm in high-performance efficiency has been made using an experimental approach Cancer (Kumar and Bavithra 2020). My study opinion is the efficient prediction of cardiovascular disease using a compressive of the Prediction of cardiovascular disease prediction to novel Logistic Regression (Shobana and Nikkath Bushra 2021).(Bhavikatti et al. 2021; Karobari et al. 2021; Shanmugam et al. 2021; Sawant et al. 2021; Muthukrishnan 2021; Preethi et al. 2021; Karthigadevi et al. 2021; Bhanu Teja et al. 2021; Veerasimman et al. 2021; Baskar et al. 2021)

The Accuracy of existing research is not properly existing in the system. The existence of the experiment is total and the improvement of accuracy of a proposed algorithm system compared to the existing model by improving. To overcome these issues a Novel Gradient boosting algorithm is implemented to improve the Prediction of cardiovascular disease using by comparing the proposed one with a novel Logistic Regression Algorithm. Now by the Above two Machine Algorithms, we have taken their own Advantages and Disadvantages in

the Current survey (Wiharto, Wiharto, Kusnanto, et al. 2017). On applying Novel Gradient boosting Algorithm Memory to the Dataset followed by Performing Observations using novel Logistic Regression and the results were plotted on a graph then there two techniques are compared based on the Result. Finally getting the best algorithm for predicting.

MATERIALS AND METHODS:

The research work was carried out in the Machine Learning laboratory lab at Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai. The sample size has been calculated using the GPower software by comparing both controllers in Supervised learning (El Boussadani et al. 2020). Two numbers of groups are selected for comparing the process and their result. In each group, 10 sets of samples and 20 samples (Dhalla et al. 2012) in total are selected for this work.

The pre-test power value is calculated using GPower 3.1 software (g power setting parameters: statistical test difference between two independent means, α=0.05, power=0.80, Two algorithms (E-RNN and novel Logistic Regression Algorithm) are implemented using Technical Analysis software. In this work, no human and animal samples were used so no ethical approval is required.

Gradient boosting Algorithm:

Novel Gradient boosting is an AI method utilized in relapse and arrangement undertakings, among others. It gives a forecast model as a troupe of feeble expectation models, which are commonly choice trees.

Pseudo Code Gradient boosting:

1. Calculate “d(x, xi)” i =1, 2, ….., n; where d denotes the Euclidean distance between the points.
2. Arrange the calculated n Euclidean distances in non-decreasing order.
83. Let k be a +ve integer, take the first k distances from this sorted list.
4. Find those k-points corresponding to these k-distances.
5. Let ki denotes the number of points belonging to the ith class among k points i.e. $k \geq 0$
6.If ki >kj ∀ I ≠ j then put x in class i.

Logistic Regression Algorithm:

Logistic Regressionis a course of demonstrating the likelihood of a discrete result given an info variable. Calculated relapse is a valuable examination technique for arrangement issues, where you are attempting to decide whether another example squeezes best into a class.

Pseudocode Logistic Regression:

1. For I←1 to k
2. For each training data instance di:
3. Set the target value for the regression to

$$Zi \leftarrow \frac{Yj - P(1|dj)}{[P(1|dj).(1-P(1|dj))]}$$

4. initialize the weight of instance Dj to P (1| DJ). (1-P(1| DJ)
5. finalize a f(j) to the data with class value (zj) & weights (WJ)
Classification Label Decision
6. Assign (class label:1) if P (1|dj) > 0.5, otherwise (class label: 2).

Statistical analysis

SPSS software is used for statistical analysis of novel approaches to efficient prediction of cardiovascular disease using novel gradient boosting compared to novel Logistic Regression with improved accuracy (Lu et al. 2022). The independent variable is novel gradient boosting accuracy and the dependent variable is efficiency. The independent T-test analyses are carried out to calculate the accuracy of the Gradient boosting for both methods.

RESULTS

Below Table shows the simulation result of the proposed Gradient boosting algorithm and the existing system novel Logistic Regression was run at different times in the google collabo with a sample size of 10. From the table, it was observed that the mean accuracy of the Machine learning Algorithms like Gradient boosting was 91.00% and the novel Logistic Regression algorithm was 92.18%The group statistics for the t-test are shown in Table 1.

The Mean, Standard Deviation, and Standard Error Mean were calculated by taking an independent variable T-test among the study groups. The Gradient boosting algorithm produces a mean difference from the novel Logistic Regression algorithm with a value of 2.09693.

Table 2 represents the Mean of the Gradient boosting algorithm which is better than the Logistic Regression algorithm with a standard deviation of 1.18200and 3.16228 respectively. From Gradient boosting algorithm and Logistic Regression algorithm in terms of mean and accuracy. The mean results, the Gradient boosting gives better accuracy than the Logistic Regression algorithm. Figure 1 gives the comparison chart of Gradient boosting the accuracy of the novel Logistic Regression algorithm is better than Logistic Regression. It is, therefore, conclusive that Gradient boosting performs better than novel Logistic Regression. The resultant plots are shown below in fig 1. The figure has been placed at the end of the paper.

DISCUSSION:

Gradient boosting and novel Logistic Regression algorithms are implemented and compared for the Prediction of cardiovascular disease Prediction to improve the accuracy by review prediction (Georgousopoulou et al. 2015). From obtained results, it is concluded that the Gradient boosting algorithm provides better accuracy results compared to the novel Logistic Regression algorithm.

In this paper, the general conversation has shown that the exhibition of various classifiers was adequate in contrast with past investigations, notwithstanding, there are without a doubt couple of impediments, such as the reliance on a particular Feature Selection procedure, for example, more dependence on relief for this situation to deliver exceptionally precise outcomes Cancer Obesity (Simmonds and Wald 2012). Moreover, a significant degree of missing values in the dataset can have an antagonistic impact. We have exhibited how to resolve the issue through the appropriate strategies (Kumar and Bavithra 2020). Accordingly, other datasets should likewise deal with this issue if the missing worth is very critical

when utilized with this model. Besides, however, our preparation dataset is sensibly broad, a bigger dataset would make the model more exact Diet will reduce Obesity is interlink to Healthcare (Nadakinamani et al. 2022). Grouping and forecasts are the fundamental methodologies of information mining. The Classification models make a difference in ordering particular, scattered information values on the other hand forecast model expected qualities that are ceaseless Cancer in Obesity. From there on utilizing the investigation result for offering web/portable applications to the clients (Sahu and Farooqui 2021). Following are the stages in the proposed approach: client enlistment and login in light of Application, dataset assortment, characterization by means of Navies Bayes, and forecast and secure information move by the method for AES Diet (Advanced Encryption Standard) and ultimately yield in PDF design. AES helps in communicating client information to the data set in a gotten way in Cancer (Manimurugan et al. 2022).

Although the results of the study are better in both experimental and statistical analysis, there are certain limitations in work. The accuracy evaluation cannot provide a better outcome on larger data sets. However, the work can be enhanced by applying optimization algorithm techniques, to achieve better accuracy. Feature selection algorithms can be used before classification to improve the accuracy of classifiers. Hence, through FS algorithms, we can reduce the computation time and improve the classification accuracy of classifiers.

CONCLUSION

From the above results and comparison, we can observe that Logistic Regression is a significantly better model identified with improved accuracy than gradient boosting.

DECLARATION

Conflict of Interests

No conflict of interest in this manuscript.

Author's Contributions

Author TVV was involved in data collection, data analysis, and manuscript writing. Author AR was involved in the conceptualization, data validation, and critical review of the manuscript.

Acknowledgments

The authors would like to express their gratitude to Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Biozone Research Technologies Pvt.Ltd, Chennai
2. Saveetha University
3. Saveetha Institute of Medical And Technical Sciences
4. Saveetha School of Engineering

REFERENCES

1. Baskar, M., R. Renuka Devi, J. Ramkumar, P. Kalyanasundaram, M. Suchithra, and B. Amutha. 2021. "Region Centric Minutiae Propagation Measure Orient Forgery Detection with Finger Print Analysis in Health Care Systems." *Neural Processing Letters*, January. https://doi.org/10.1007/s11063-020-10407-4.
2. Bhanu Teja, N., Yuvarajan Devarajan, Ruby Mishra, S. Sivasaravanan, and D. Thanikaivel Murugan. 2021. "Detailed Analysis on Sterculia Foetida Kernel Oil as Renewable Fuel in Compression Ignition Engine." *Biomass Conversion and Biorefinery*, February. https://doi.org/10.1007/s13399-021-01328-w.
3. Bhavikatti, Shaeesta Khaleelahmed, Mohmed Isaqali Karobari, Siti Lailatul Akmar Zainuddin, Anand Marya, Sameer J. Nadaf, Vijay J. Sawant, Sandeep B. Patil, Adith Venugopal, Pietro Messina, and Giuseppe Alessandro Scardina. 2021. "Investigating the Antioxidant and Cytocompatibility of Mimusops Elengi Linn Extract over Human Gingival Fibroblast Cells." *International Journal of Environmental Research and Public Health* 18 (13). https://doi.org/10.3390/ijerph18137162.
4. B., Shiva Shanta Mani, and V. M. Manikandan. 2021. "Heart Disease Prediction Using Machine Learning." *Handbook of Research on Disease Prediction Through Data Analytics and Machine Learning*. https://doi.org/10.4018/978-1-7998-2742-9.ch018.
5. Dhalla, Naranjan S., Heinz Rupp, Aubie Angel, and Grant N. Pierce. 2012. *Pathophysiology of Cardiovascular Disease*. Springer Science & Business Media.
6. El Boussadani, B., C. Benajiba, A. Aajal, A. Ait Brik, O. Ammour, J. El Hangouch, O. Oussama, B. Oussama, N. Tahiri, and Z. Raissuni. 2020. "[COVID-19 pandemia: Impact on the cariovascular system. Data of 1 April 2020]." *Annales de cardiologie et d'angeiologie* 69 (3): 107–14.
7. Georgousopoulou, Ekavi N., Demosthenes B. Panagiotakos, Christos Pitsavos, Christodoulos Stefanadis, and ATTICA study group. 2015. "Assessment of Diet Quality Improves the Classification Ability of Cardiovascular Risk Score in Predicting Future Events: The 10-Year Follow-up of the ATTICA Study (2002-2012)." *European Journal of Preventive Cardiology* 22 (11): 1488–98.
8. Govoni, Stefano, Pierluigi Politi, and Emilio Vanoli. 2020. *Brain and Heart Dynamics*. Springer.
9. Karobari, Mohmed Isaqali, Syed Nahid Basheer, Fazlur Rahman Sayed, Sufiyan Shaikh, Muhammad Atif Saleem Agwan, Anand Marya, Pietro Messina, and Giuseppe Alessandro Scardina. 2021. "An In Vitro Stereomicroscopic Evaluation of Bioactivity between Neo MTA Plus, Pro Root MTA, BIODENTINE & Glass Ionomer Cement Using Dye Penetration Method." *Materials* 14 (12). https://doi.org/10.3390/ma14123159.
10. Karthigadevi, Guruviah, Sivasubramanian Manikandan, Natchimuthu Karmegam, Ramasamy Subbaiya, Sivasankaran

Chozhavendhan, Balasubramani Ravindran, Soon Woong Chang, and Mukesh Kumar Awasthi. 2021. "Chemico-Nanotreatment Methods for the Removal of Persistent Organic Pollutants and Xenobiotics in Water - A Review." *Bioresource Technology* 324 (March): 124678.

11. Kumar, Digvijay, and Bavithra. 2020. "Cardiovascular Disease Prediction Using Machine Learning." *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*. https://doi.org/10.32628/cseit20659.
12. Lu, Haoxuan, Yudong Yao, Li Wang, Jianing Yan, Shuangshuang Tu, Yanqing Xie, and Wenming He. 2022. "Research Progress of Machine Learning and Deep Learning in Intelligent Diagnosis of the Coronary Atherosclerotic Heart Disease." *Computational and Mathematical Methods in Medicine* 2022 (April): 3016532.
13. Mackay, Judith, George A. Mensah, Shanthi Mendis, and Kurt Greenlund. 2004. *The Atlas of Heart Disease and Stroke*.
14. Manimurugan, S., Saad Almutairi, Majed Mohammed Aborokbah, C. Narmatha, Subramaniam Ganesan, Naveen Chilamkurti, Riyadh A. Alzaheb, and Hani Almoamari. 2022. "Two-Stage Classification Model for the Prediction of Heart Disease Using IoMT and Artificial Intelligence." *Sensors* 22 (2). https://doi.org/10.3390/s22020476.
15. Mehta, Jawahar L., and Jean McSweeney. 2018. *Gender Differences in the Pathogenesis and Management of Heart Disease*. Springer.
16. Mendis, Shanthi, Pekka Puska, Bo Norrving, and World Health Organization. 2011. *Global Atlas on Cardiovascular Disease Prevention and Control*. World Health Organization.
17. Muthukrishnan, Lakshmipathy. 2021. "Nanotechnology for Cleaner Leather Production: A Review." *Environmental Chemistry Letters* 19 (3): 2527–49.
18. Nadakinamani, Rajkumar Gangappa, A. Reyana, Sandeep Kautish, A. S. Vibith, Yogita Gupta, Sayed F. Abdelwahab, and Ali Wagdy Mohamed. 2022. "Clinical Data Analysis for Prediction of Cardiovascular Disease Using Machine Learning Techniques." *Computational Intelligence and Neuroscience* 2022 (January): 2973324.
19. Nesaragi, Naimahmed, Ashish Sharma, Shivnarayan Patidar, and U. Rajendra Acharya. 2022. "Automated Diagnosis of Coronary Artery Disease Using Scalogram-Based Tensor Decomposition with Heart Rate Signals." *Medical Engineering & Physics*, April, 103811.
20. Preethi, K. Auxzilia, K. Auxzilia Preethi, Ganesh Lakshmanan, and Durairaj Sekar. 2021. "Antagomir Technology in the Treatment of Different Types of Cancer." *Epigenomics*. https://doi.org/10.2217/epi-2020-0439.
21. Sahay, Sandeep, Jiken Bhatt, Sarah Beshay, Ashrith Guha, Duc T. Nguyen, Edward A. Graviss, and Sherif F. Nagueh. 2022. "E-REVEAL Lite 2.0 Scoring for Early Prediction of Disease Progression in Pulmonary Arterial Hypertension." *Pulmonary Circulation* 12 (1): e12026.

22. Sahu, Sachin, and Zuber Farooqui. 2021. "High Accurancy and Low Risk Prediction and Diagnosis Heart Disease Using Gradient Boosting Algorithm." *International Journal of Computer Applications*. https://doi.org/10.5120/ijca2021921201.
23. Sawant, Kashmira, Ajinkya M. Pawar, Kulvinder Singh Banga, Ricardo Machado, Mohmed Isaqali Karobari, Anand Marya, Pietro Messina, and Giuseppe Alessandro Scardina. 2021. "Dentinal Microcracks after Root Canal Instrumentation Using Instruments Manufactured with Different NiTi Alloys and the SAF System: A Systematic Review." *NATO Advanced Science Institutes Series E: Applied Sciences* 11 (11): 4984.
24. Shanmugam, Vigneshwaran, Rhoda Afriyie Mensah, Michael Försth, Gabriel Sas, Ágoston Restás, Cyrus Addy, Qiang Xu, et al. 2021. "Circular Economy in Biocomposite Development: State-of-the-Art, Challenges and Emerging Trends." *Composites Part C: Open Access* 5 (July): 100138.
25. Shobana, G., and S. Nikkath Bushra. 2021. "Prediction of Cardiovascular Disease Using Multiple Machine Learning Platforms." *2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*. https://doi.org/10.1109/icses52305.2021.9633797.
26. Simmonds, Mark C., and Nicholas J. Wald. 2012. "Risk Estimation versus Screening Performance: A Comparison of Six Risk Algorithms for Cardiovascular Disease." *Journal of Medical Screening* 19 (4): 201–5.
27. Tavares, João Manuel R., Satyajit Chakrabarti, Abhishek Bhattacharya, and Sujata Ghatak. 2021. *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2020, Volume 3*. Springer Nature.
28. Trevisan, Caterina, Giuseppe Sergi, and Stefania Maggi. 2020. "Gender Differences in Brain-Heart Connection." *Brain and Heart Dynamics*. https://doi.org/10.1007/978-3-030-28008-6_61.
29. Veerasimman, Arumugaprabu, Vigneshwaran Shanmugam, Sundarakannan Rajendran, Deepak Joel Johnson, Ajith Subbiah, John Koilpichai, and Uthayakumar Marimuthu. 2021. "Thermal Properties of Natural Fiber Sisal Based Hybrid Composites – A Brief Review." *Journal of Natural Fibers*, January, 1–11.
30. Wiharto, Wiharto Wiharto, Harianto Herianto, and Hari Kusnanto. 2017. "The Analysis of Performace Model Tiered Artificial Neural Network for Assessment of Coronary Heart Disease." *International Journal of Electrical and Computer Engineering (IJECE)*. https://doi.org/10.11591/ijece.v7i4.pp2183-2191.
31. Wiharto, Wiharto Wiharto, Hari Kusnanto, and Herianto Herianto. 2017. "Hybrid System of Tiered Multivariate Analysis and Artificial Neural Network for Coronary Heart Disease Diagnosis." *International Journal of Electrical and Computer Engineering (IJECE)*. https://doi.org/10.11591/ijece.v7i2.pp1023-1031.

32. Yadav, Dhyan Chandra, and Saurabh Pal. 2021. "Analysis of Heart Disease Using Parallel and Sequential Ensemble Methods With Feature Selection Techniques." *International Journal of Big Data and Analytics in Healthcare*. https://doi.org/10.4018/ijbdah.20210101.oa4.

Tables and Figures

Table 1. The LR(92.1820) method and grouped statistics were compared using group statistics for recorded data from simulation for 10 iterations (91.00).

Algorithm	Algorithm	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	LR	10	92.1820	3.46203	1.54826
	GB	10	91.0000	3.16228	1.41421

Tabel 2: Independent sample T-test is applied for the dataset fixing confidence as 95% and level of significance value p=0.001(P< 0.05)

Accuracy		Levene's Test for Equality of Variance		T-test for Equality of Means						
		f	Sig	t	df	Sig.(2-tailed)	Mean Difference	Std.Error Difference	95% Confidence of the Differences	
									Lower	Upper
ACCURACY	Equal variances assumed	.038	.850	.564	8	<.001	1.18200	2.09693	-3.65353	6.01753
	Equal variances not assumed			.564	7.935	<.006	1.18200	2.09693	-3.66041	6.02441

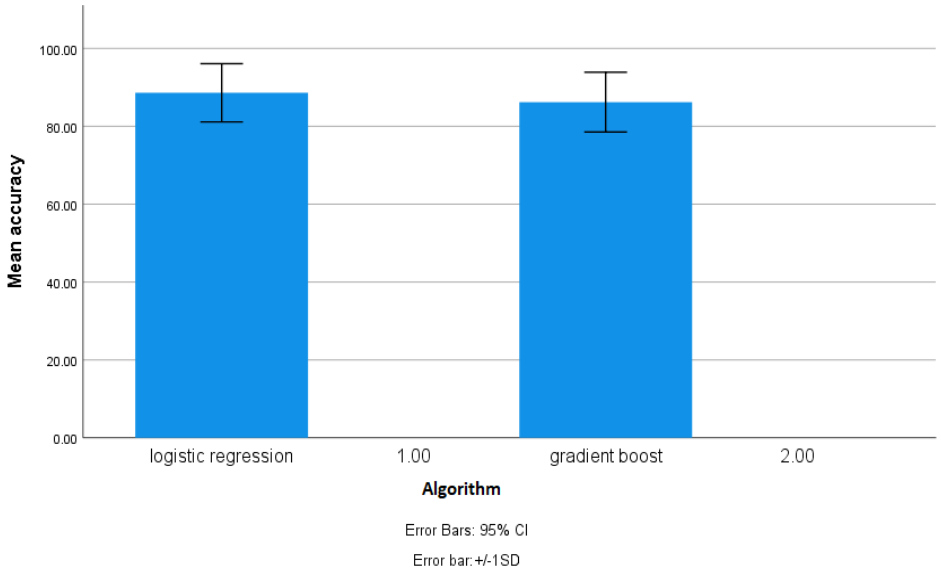

Fig. 1. Comparison of Logistic Regression and Gradient boost analysis concerning mean exactness. The mean exactness of Logistic Regression is better compared to Gradient boost. X-Axis: Logistic Regression and Gradient boost, Y-Axis: Mean exactness of location ± 2 SD.