SmartHeart RF: A Machine Learning Approach to Cardiovascular Disease Detection in Obesity

Abhishek Khanna[1], Sapna[2], Arjun Puri[3]

{Abhishek.khanna@chitkara.edu.in[1], sapnakcp@gmail.com[2] , arjun puri@chitkara.edu.in[3]}

Assistant Professor, Department of Computer Science and Engineering, Chitkara University, Rajpura, Patiala[1] ,

Assistant Professor, School of Commerce and Management, GSSDGS Khalsa College, Patiala [2],

Assistant Professor, Department of Computer Science and Engineering, Chitkara University, Rajpura, Patiala[3]

Abstract: Cardiovascular disease (CVD) is one of the leading causes of death worldwide, with obesity being a major contributing factor across all age groups. Addressing obesity-related CVD risk requires a comprehensive approach that evaluates multiple interconnected factors using advanced predictive techniques. This study focuses on identifying key variables involved in assessing CVD risk linked to obesity, including biomarkers, anthropometric measurements, smoking habits, medical history, hypertension, and diabetes. These factors are integrated into established risk assessment tools, such as Random Forest–based machine learning models, along with modern statistical methods to improve prediction accuracy for individuals aged 10 years and above. Additionally, the use of cardiac imaging datasets provides deeper insights into heart structure and function, enhancing traditional risk assessment strategies. The study also highlights the importance of lifestyle factors such as diet, physical activity, and stress in shaping CVD risk profiles. Long-term monitoring is emphasized as a critical component for tracking changes and enabling early interventions. Overall, the integrated framework supports more accurate prediction and effective management of obesity-related CVD risk, contributing to improved cardiovascular health outcomes.

Keywords: Cardiovascular disease (CVD), obesity, Random Forest Machine Learning, age, height, weight, ap_high, and ap_low.

1 Introduction

Over the past few years, there has been a significant increase among people suffering from obesity disease. Obesity within an individual indicates a negative influence on many physiological processes and enhances the risk of Cardiovascular disease (CVD) [2,3]. Globally, obesity is recognized as a significant public health concern that has a direct impact not only on individual's health; but also on entire healthcare system [4]. The precise nature obesity having a direct or significant impact on risk of developing CVD is relatively unclear. Hence, there

IICTDS 2025, December 19-20, Chandigarh, India
Copyright © 2026 EAI
DOI 10.4108/eai.19-12-2025.2362620

arises a necessity for precise risk prediction of individuals with CVD with different degrees of obesity is one of the major cause of concern which this manuscript aims to solve. Although the link between obesity and CVD is well-established, little is known about the precise components of obesity that raise the risk of CVD. To better assess CVD risk and guide preventative strategies and interventions, it is imperative to propose prediction model that encompass a wide variety of obesity-related characteristics. Related studies conducted during the past few years have revealed significant impact of obesity towards CVD [5]. However, little is known about the exact roles that obesity-related variables include family history, body mass index (BMI), and lifestyle choices play in predicting CVD risk. Hence there is a dire need of a systematic study that can identify or co-relate the significant impact of obesity on CVD.

The findings of the above said parameters can further be used for investigation purpose to apply machine learning methods with an aim to predict CVD. Our hypothesis is that by integrating lifestyle habits, family history, and Body Mass Index (BMI) into predictive models, CVD risk assessment will be more accurate than it would be with conventional techniques. The next section highlights some of the most prominent studies conducted in this context.

2 Literature Review

Before performing experimentation on the acquired dataset, it is very much necessary to first evaluate the current standings for the existing study. Hence an extensive literature review has been done in order to evaluate the methodologies that have already been incorporated for the study under consideration along with their acquired results. The selection of reserach articles for the purpose of literature review has been done on the basis of highest number of citations over the past few years.

In a study conducted by. Lavie et al. (2009) highlighted towards an intricative relationship among the two prominent variables, i.e. , obesity and cardiovascular disease. Furthermore, their study aimed to determined a divisional identification factors among the most prominent factors, that were inclned towards indicating higher significant risk factor and also served as a risk marker for cardiovascular events. By delving into the underlying mechanisms linking obesity to heart disease, the authors underscored the importance of targeted interventions to address obesity-related cardiovascular risk factors effectively [6].

Aekplakorn et al. (2006) in their study highlighted an extensive review investigating the impact of obesity on cardiovascular disease. Furthermore, the study also highlighted impact of various prospectives such as pathophysiological, epidemiological, and clinical. All the attributes stated above highlighting a profound impact of obesity on the development and progression of cardiovascular conditions such as dyslipidaemia, hypertension, coronary artery disease, and heart failure. Lastly, study also focused towards a dire urgency for identification of other comprehensive strategies to address the ever increasing impact of obesity-related heart disease [7].

Chandrasekaran et al. (2024) in their findings highlighted a significant interdependency of cardiovascular disease due to obesity. This comprehensive review delived into the pathophysiology of obesity-related heart disease, including insulin resistance, inflammation, and endothelial dysfunction. The authors of the present study further highlighted the evaluation and management of obesity-related cardiovascular risk factors, stressing the importance of lifestyle modifications, and weight loss interventions in mitigating probable cardiovascular risk among obese individuals [8].

Kenchaiah et al. (2002) in their findings highlighted the significant impact of obesity resuting in heart faliure. Their study was printed in the Journal of Medicine in New England in 2002,

titled "Obesity and the Risk of Heart Failure." This seminal research investigated on all paramters assocciated betwen obesity and heart failure. Furthermore, the study utilized a large community-based cohort and found a strong and graded connection between increasing Body Mass Index (BMI) and the risk of heart failure, even with additional cardiovascular risk variables when taken into consideration. Their findings explicitly highlighted the critical importance of obesity prevention and management stratergies that helped in reducing the burden of heart failure and improving towards cardiovascular health outcomes [9].
The next section of the manuscirpt highlights the objective of the study.

3 Objective of the study

The main objective of this manuscript is to evaluate and predict the possibility of CVD based on the associated paramters of obesity. This objectve has been attained by evaluating various paramters over a python classification technique, i.e. ,Random Forest. By incorporating the proposed technique, the researchers of the present study were able to attain the following objectives:
Develop Accurate Predictive Model: Build a machine learning model, that is capable of accurately predicting cardiovascular disease risk in individuals affected by obesity, leveraging Python classification algorithms to achieve high predictive performance.
Identify the key predictors: By considering and evaluating paramters such as genetic, demographic, clinical, and lifestyle factors, the researchers were able to determine that obesity had a direct and significant impact on cardiovascular disease risk. Furthermore, by incorporating feature selection technique the resaerchers were able to improve the clarity and clinical usefulness of proposed model within the study.
Evaluate Model Generalization: Increase how understandable the predictive models are across different groups and environments by thoroughly validating them to ensure dependable performance in actual clinical situations, not just on the training set.
Enhance Model Interpretability: In addition, the researchers of the present study were able to idenitfy the precise interpretability of predictive models by deploying Explainable Artificial Intelligence (XAI) techniques the helped in enabling healthcare practitioners to understand and build a confidence towards a sense of trust among the proposed model predictions taken into consideration for the present study and also by fostering shared decision-making and personalized interventions for the final outcomes.
Translate Research into Practice: The final objective of the present study is to bridge the gap between existing clinical practices and research by proposing predictive models into an actionable insights, that actis as decision support tools and facilitates early CVD detection, risk stratification, and provides accurate preventive interventions for cardiovascular diseases that are caused due to obesity.
The next section of the study highlights the methodology adopted to conduct the study.

4 Methodology

Based on existing studies conducted by various esteemed researchers, Random Forest Machine Learning techique was taken into consideration. Random Forest Machine Learing technique is one of the oldest and most frequently used method of group learning for tasks involving regression and classification [10]. During training, it builds several decision trees, from which it extracts the mode of the classes (for classification) or the mean forecast (for regression) of each tree separately. Random Forest introduces randomness during the construction of each tree, including:

Random selection of features at each node split.
Bootstrapping of the trained data.
An another reason to choose Random Forest Machine Learing is because this technique is renowned for its resilience to overfitting, robustness, and capacity to handle high-dimensional datasets [11]. To identify the adverse effects of obesity on CVD, various parameters have been taken into consideartion, such as age, height, weight, gender, Systolic blood pressure (ap_high), Diastolic blood pressure (ap_low), cholestrol, gulucose, smoking and drinking habbits. An alogrithm was designed initially so as to understand the evaluation process in advance.

Algorithm
Read the obesity dataset.
Pre-process the data to calcualte the body mass index and obesity levels categorized using predefined BMI thresholds.
Explore data to analyse the dataset characteristics such as numerical value distribution and correlations.
Use histogram, pairplot, correlation matrix to visualize the datasets characteristics.
Split data into training and testing sets.
Train Random Forest Classifier.
Trained model then makes predictions on the acquired dataset.

In order to verify the outcomes of the proposed model, Cross Validation has also been performed on the outcomes. Cross-validation is an important mechanism that aims on monitoring the effectiveness of a machine learning model. The collection is split up into several subsets, or folds, and these subsets are iteratively used for training and validation. A dataset is split into k equal-sized folds for k-fold cross- validation. The model is trained on k-1 folds and verified on the remaining fold. A general assessment of The performance metrics are aggregated to judge the model's performance (such as accuracy) over the course of the k iterations of this process, which is performed k times with a new fold for validation each time. In contrast to a single split of the test as well as train, cross-validation yields a more accurate estimate of performance and aids in evaluating how effectively a model generalises to unknown data.
Confusion Matrix: A confusion matrix is a table that is being used to examine the effectiveness of a classification model. It provides a brief description of how many predictions the model is made on a classification task were accurate and inaccurate.
There are four primary parts to the confusion matrix:
True Positive (TP): The positive instances that are accurately suggested.
True Negative (TN): The negative instances that are accurately suggested.
False Positive (FP): The positive instances that are inaccurately estimated.
False Negative (FN): The negative instances that are inaccurately estimated.
The confusion matrix's components are used to compute a number of performance measures, including F1-score, recall, accuracy, and precision.
In order to have a precise evaluation outcome, Feature selection technique was also incorporated while choosing the parameters. The identification of a subset of extracted features (variables) from a broader set of features is known as feature selection. By concentrating on the most informative features, it seeks to increase interpretability, decrease overfitting, and improve model performance. Three primary categories can be used to categorise feature selection techniques:
Filter methods: Consider quantitative measurements or connections with the target variable when assessing the significance of features.

Wrapper methods: Analyse feature subgroups by training and testing the model with various feature combinations.
Embedded methods: Include feature selection in the process of training the model. (e.g., significance of features in tree-based models like Random Forest). Feature selection aids in decreasing the dataset's dimensionality and improving the efficiency of the learning algorithm.
The proceeding section of the study highlights the results for the study.

5 Results and Analysis

The study's findings indicate a significant association between the risk of cardiovascular illnesses and factors connected to obesity. When the dataset was trained on the proposed machine learning technique on the basis of identified features, the the proposed modelling technique outperformed conventional methods of predicting the risk of CVD. Based on the results, the conclusions indicated that dietary factors, smoking habits, family history, and BMI were all high indicators of high risk of CVD. High sensitivity and specificity are demonstrated by the machine learning model in identifying people who are more likely to acquire cardiovascular illnesses.

Negative findings are also reported, such as the lack of a meaningful correlation between a few lifestyle factors and the risk of CVD. These results add to our comprehensive understanding of the intricate connection between cardiovascular health and obesity.

The strongest correlations seem to involve demographic variables (age, height, weight) to lifestyle variables (ap_lo, ap_hi). Lifestyle choices, particularly diet (vegetable consumption) and physical activity, appear to have a more significant correlation with obesity levels. Figure 1 depicts co-relation coefficients for various parameters taken into consideration.

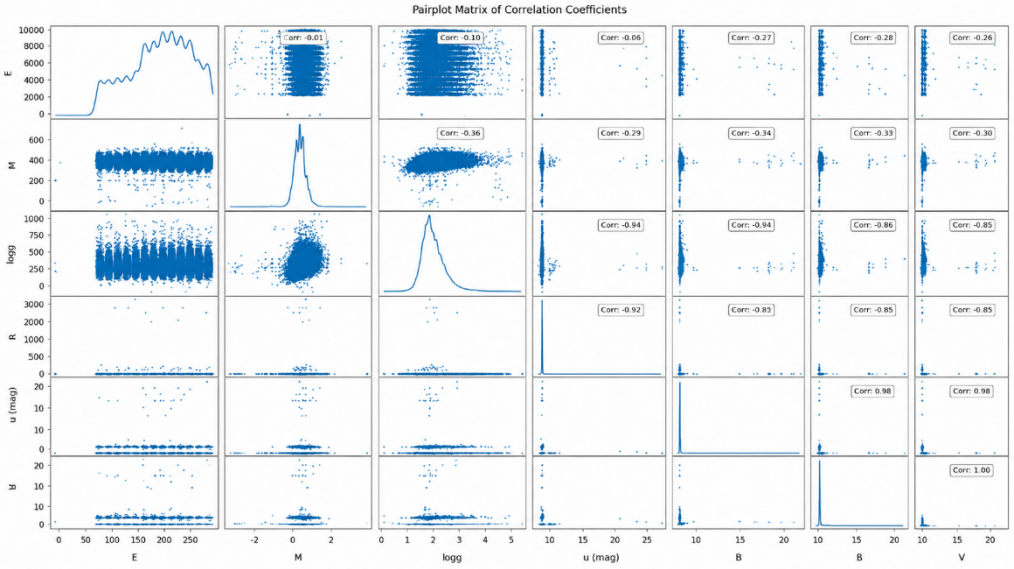


Fig. 1 Co-relation coefficients for various parameters taken into consideration for the present study

Figure 2 reflects the Boxplots for 5 prominent varilables *i.e.*, *age*, *height*, *weight*, *ap_hi*, and *ap_lo* have been identified based on the findings that have most significant impact among people who already are obesed and further have possibility of cardio vascular disease.

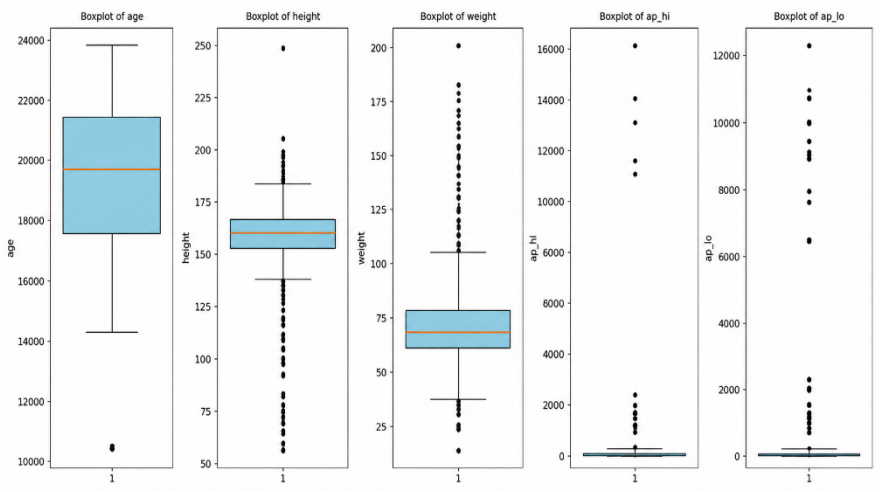


Fig. 2 Boxplots for 5 prominent varilables

Furthermore, it has been conculed that adding obesity related variables to predictive models impove the precision of estimating the risk of CVD. Important factors while determining CVD risk include lifestyle choices, family history and body mass index.

The significance and originality of the study's conclustions are highlighted by comparison with the body of existing literature. The proposed models predictive surpasses that of conventional risk assessment techniques, underscoring the need of taking into account a varieyt of factors when predicting the variables. Figure 3 reflects the corelation matrix for various attributes taken into considearation for the present study.

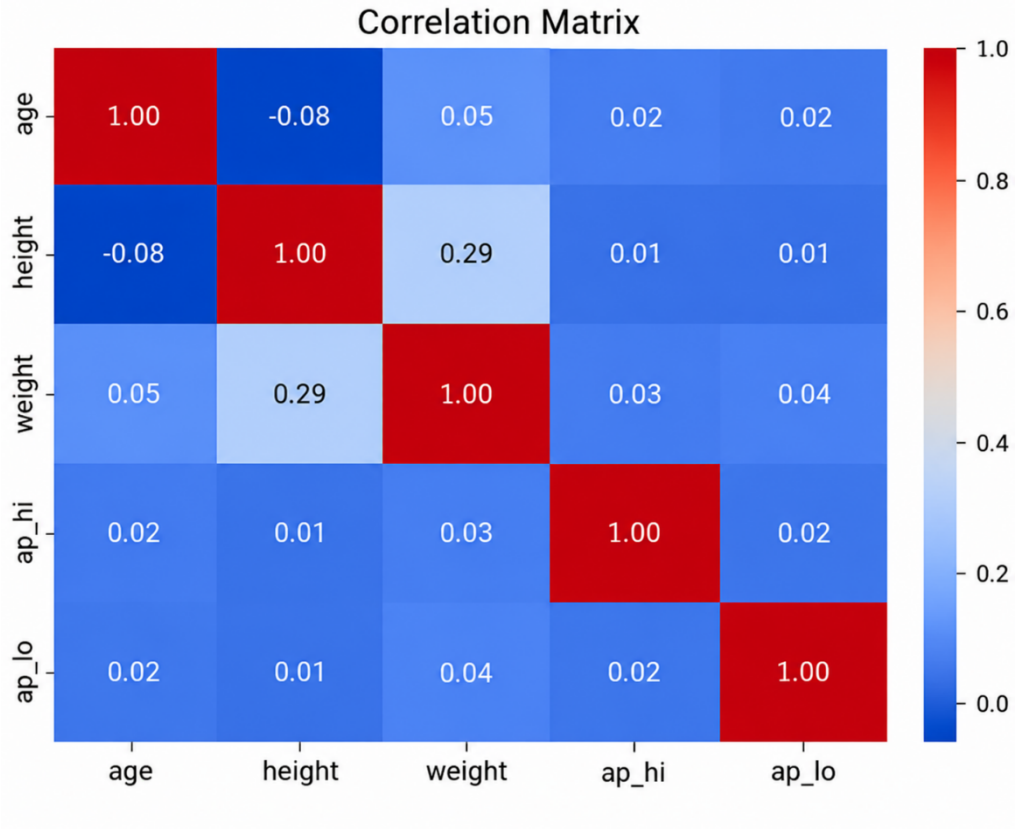


Fig. 3 Corelation matrix driven among various parameters.

The study's limitations include its dependence on self-reported data for lifestyle and adding related characterstics, the need for validation in larger, more diverse cohorts, and possible confounding variables that were overlooked in the analysis.

The results have consequences for public health initiatives that target modifiable risk factors incuding obesity and bad life style choices in an effort to prevent cardivascular illness in addition to the variables taken into considearation for the present study.

The final section of the present study highlights the conclusion and future scope for the present study.

6 Conclusion and Future Scope

To conclude, this study exaplifies as to how machine learning techniques can be incorporated to determine cardiovascular disease risk using obesity related parameters. The predictive model proposed within the present study has been trained to evaluate the parameters and to depict the significant impact of the associated parameters of obesity can improve in tracking and predicting the accuracy of risk estimation, enabling more targeted as well as timely interventions with an aim to avoid CVD probability. The findings furthermore deepen our understanding on complex relationship among obesity and various cardiovascular conditions, that may arise from time to

time. The study also provides an an indepth understanding of personalized risk assessment and effective disease management and prevention. Lastly, the study under consideration also aims at recomending early indicators of CVD by timely idenitfying them and later can be addressed through the following proactive measures:

- Leveraging Wearable Devices and Health Sensors: Embrace the integration of wearable devices and health sensors to capture real-time physiological data, enabling continuous monitoring of cardiovascular health indicators and facilitating dynamic risk assessment in individuals affected by obesity.
- Enhanced Model Interpretability: Current findings of the present study can further be invested for research purposes with an aim to enhance the interpretability of predictive models by develping Explainable Artificial Intelligence (XAI) techniques, that aims to enhance the knowledge of healthcare practitioners as well as ailing patients to better understand the underlying risk factors.
- Population-specific Risk Assessment: Another aspect can be by tailoring predictive models among specific population groups by taking into account elements like their socio-economic standings and geographic location so as to account for disparities of cardiovascular disease risk and ensure that they are pvovided with equitable healthcare delivery.
- Clinical Translation and Implementation: The final aspect of the present study is to bridge the gap among research and clinical practitioners by facilitating them with translation and implementation of predictive models into modern day healthcare privilages by integrating acquired data and results over Electronic Health Records (EHRS) and Decision Support Systems (DSS) with an aim to be proactive in terms of providing timely personalized care interventions.

Looking ahead for enhancing this work for further basis, our work aims to provide re-shaping healthcare practices and aims at providing vital collabrations, determining innovative techniques, and ethical responsibility to determine harness potential abilities of predictive analytics to provide proactive care, by identifying unique needs of individual affected by obesity and cardiovascular disease.

References

[1] N. Sinha, T. Jangid, A. M. Joshi, and S. P. Mohanty, “iCARDO: A machine learning based smart healthcare framework for cardiovascular disease prediction,” arXiv preprint arXiv:2212.08022, (2022).
[2] A. Elagizi, S. Kachur, S. Carbone, C. J. Lavie, and S. N. Blair, “A review of obesity, physical activity, and cardiovascular disease,” Current Obesity Reports, vol. 9, no. 4, pp. 571–581, (2020).
[3] R. M. Krauss, M. Winston, B. J. Fletcher, and S. M. Grundy, “Obesity: Impact on cardiovascular disease,” Circulation, vol. 98, no. 14, pp. 1472–1476, (1998).

[4] D. Mohajan and H. K. Mohajan, “Obesity and its related diseases: A new escalating alarming in global health,” Journal of Innovations in Medical Research, vol. 2, no. 3, pp. 12–23, (2023).
[5] G. Battineni et al., “Impact of obesity-induced inflammation on cardiovascular diseases,” International Journal of Molecular Sciences, vol. 22, no. 9, Art. no. 4798, (2021).
[6] C. J. Lavie, R. V. Milani, and H. O. Ventura, “Obesity and cardiovascular disease: Risk factor, paradox, and impact of weight loss,” Journal of the American College of Cardiology, vol. 53, no. 21, pp. 1925–1932, (2009).
[7] W. Aekplakorn, V. Kosulwat, and P. Suriyawongpaisal, “Obesity indices and cardiovascular risk factors in Thai adults,” International Journal of Obesity, vol. 30, no. 12, pp. 1782–1790, (2006).
[8] P. Chandrasekaran and R. Weiskirchen, “The signaling pathways in obesity-related complications,” Journal of Cell Communication and Signaling, vol. 18, no. 2, Art. no. e12039, (2024).
[9] S. Kenchaiah et al., “Obesity and the risk of heart failure,” New England Journal of Medicine, vol. 347, no. 5, pp. 305–313, (2002).
[10] M. R. Segal, “Machine learning benchmarks and random forest regression,” (2004).
[11] A. Wilson and M. R. Anwar, “The future of adaptive machine learning algorithms in high-dimensional data processing,” International Transactions on Artificial Intelligence, vol. 3, no. 1, pp. 97–107, (2024).