Grenze International Journal of Engineering and Technology, June Issue

Gender-Specific Predictive Modeling of Bidirectional Risk between Heart Disease and Diabetes Mellitus using Machine Learning

Athul Narayan. G[1], Binayak Dutta[2] and Somnath Sinha[3]
[1-3]Christ University, Bangalore, India
Email: athul.g@bsccsh.christuniversity.in, binayak.dutta@christuniversity.in, Somnath.sinha@christuniversity.in

***Abstract*— Cardiovascular diseases (CVD) and diabetes are two of the most challenging global health problems and often coexist to escalate morbidity and mortality. Early prediction of these diseases using machine learning makes way for early intervention itself. In this paper, the authors provide a comparative study of the Logistic Regression method and the Random Forest approach used to predict two of the most challenging intersections of clinical diseases: (1) the possibility of developing cardiovascular diseases when already suffering from diabetes and vice versa, and (2) the possibility of developing diabetes when already affected with cardiovascular diseases. In this work, the structured data has been used to segregate the community of people into four groups: Everyone, Pregnant Women, Women, and Men. The results indicate the possibility of large differences in probabilities of predicting diseases across the sexes and also across the group of pregnant and non- pregnant women and men. The results also support the fact that the Logistic Regression approach outperforms the Random Forest approach in predicting the possibility of developing cardiovascular diseases across almost all groups. This suggests the applicability of the Logistic Regression approach when the predicting factors are mostly linear in nature. However, the performance of the two models was found equally efficient in predicting the possibility of developing diabetes. This suggests that the effectiveness of the predicting model might depend upon the characteristics of the particular disease being studied.**

***Index Terms*— machine learning, cardiovascular diseases, diabetes mellitus, sex-specific prediction, logistic regression, random forest.**

I. INTRODUCTION

Chronic illnesses – especially cardiovascular diseases (CVDs) and diabetes mellitus (DM) – remain a serious global health problem. According to the World Health Organization, cardiovascular diseases are the cause of almost 18 million deaths each year and are the leading cause of mortality glob- ally. A global health burden of diabetes has been estimated to be in the order of hundreds of millions of people worldwide. Besides being the cause of complications from diabetes, it also substantially increases the likeliness of developing cardiovascular diseases due to the interlinkages of the two conditions – diabetes and cardiovascular diseases. In recent years, the role of Machine Learning (ML) has been emerging as a significant approach to model the progression of diseases and identify latent patterns in clinical data. ML risk prediction models can be used to aid medical professionals in decision- making and improve the efficiency of screening.

Grenze ID: 01.GIJET.12.2.69_24

However, many conventional and advanced models fail to account for various vital differences. Sex differences in the progression of diseases due to hormonal factors can often lead to differences in the risk factors of men and women. In addition to the differences mentioned above, there are also differences in the cardiovascular, hormonal, and metabolic factors of pregnant women. These differences can often impact the development of diseases and can be missed in ML-based models of research.
To fill this knowledge gap, this study proposes the development and testing of four different models: (1) Every-one, (2) Pregnant Women, (3) Women (non-pregnant), and (4) Men. Two popular classifiers—Logistic Regression (LR) and Random Forest (RF)—are applied to measure the capacity of the model to predict the existence of: i. Heartdisease in diabetic patients, and ii. The existence of diabetes in patients suffering from pre-existing heart diseases. This study moves beyond the binary classification model and results in the development of clinically interpretable risk scores through predicted probabilities. This will aid in the improvement of risk stratification and unmask sex- and pregnancy-related inequities in the development of clinical models predicting the likelihood of diabetes and heart disease in order to provide equitable models that predict the likelihood of the ailment.

II. Literature Review

A. Introduction to Machine Learning in Healthcare

Machine learning has developed the way predictions about diseases, as well as risks, are made. Beam and Kohane [3] proved how leveraging big data and machine learning can aid doctors in making more informed decisions through patterns in large medical data sets. Kohane [3] demonstrated that through the utilization of big data and machine learning, it has become possible to make improved decisions in health- care. Rajkomar et al. [10] observed: "Whereas machine learning in medicine has matured from theory to practice, it has been shown to make more accurate predictions about patient out- comes than conventional statistical approaches." Rajkomar et al. [10] demonstrated the applications of machine learning in medicine and stressed that machine learning produced better than conventional ones in predicting outcomes. With the use of electronic health records, advanced computer power, as well as sophisticated algorithms, predictive models can now address highly complex questions. Electronic health records, computer processing capacity, and complex algorithms has created a situation where predictive modeling can address complex clinical questions.

B. Diabetes Prediction Using Machine Learning

Diabetes prediction has also become an important target in machine learning applications. In a review performed by Kavakiotis et al. [15], it was found that the ensemble methods and support vector machines were, in fact, more accurate than the conventional statistical models when applied to heterogeneous clinical data. Feature selection approaches and algorithms were found to possess significant "influence on the accuracy of the model," according to Zou et al. [12], who emphasized the need for incorporating genetic markers and conventional clinical variables.
Huo and Xu [8] have specifically analyzed logistic regression, Decision Trees, Random Forest, and Neural Net-works, and finding that although complex models performed marginally better in accuracy, simpler interpretable models like logistic regression were rendered more clinically useful through their transparency and ease of implementation in clinical settings. The American Diabetes Association [2] gives today's guidelines with regards to the provision of medical care based on evidence, highlighting the critical need for evidence-based predictive tools in diabetes care.

C. Cardiovascular Disease Prediction

Prediction of cardiovascular disease cases is another key application domain for machine learning in healthcare. Al- Ghamdi and Al-Amri [1] authored an article that particularly targeted comparative analysis of machine learning models for early prediction of risk of cardiovascular diseases in diabetic patients, focusing on the important clinical question of cardiovascular complications especially in an already high- risk group. Their study found that diabetic patients display specific patterns of risk factors demanding predictive models specific to the situation at hand rather than the general population algorithms.
In their study, Dinh et al. employed a data-driven method to predict both diabetes and cardiovascular disease at the same time, recognizing the bidirectional relationship and shared pathophysiology among these diseases. Their integrated modeling framework proved it was possible for joint models of prediction to improve the identification of patients at risk of both conditions rather than separate disease-specific models.

D. Logistic Regression in Medical Prediction

Logistic Regression is still an essential methodology in medical predictions due to its interpretability and medical acceptability. Firth [6] had significant work in demonstrating the robustness of performance for logistic regression in comparison to bias reduction techniques in maximum likelihood estimation. This is particularly relevant in instances where sample sizes and rare events are common in medical data. In [5] comparing Logistic Regression and Random Forest, in terms of comparing between Logistic Regression and Random Forest for tasks that involve predicting in the field of healthcare, it was found that logistic regression was executed competently in competition, despite its simplicity, particularly "when relationships between predictors and outcomes were examined, largely linear." This is particularly significant in clinical practice due to its interpretability, as it faces fewer barriers in terms of implementation and is thus more acceptable in healthcare.

E. Random Forest and Ensemble Methods

Random Forest has gained recognition in the medical prediction domain due to the potential of handling complex non-linear relationships without any need for complex feature engineering. In this regard, the Random Forest algorithm was proposed by Breiman [4] , who provided a theory and proof of the algorithm's resistance to overfitting through ensemble averaging. This paper was a landmark and paved the way for the widespread acceptance of the algorithm for various domains, including healthcare analytics. Chen et al. [11] proposed the potential of predicting diseases through machine learning, where large volumes of data from the health community are used. The study proved that ensemble learning effectively combines information from different sources, such as clinical reports, laboratory results, and patient outcomes, and identified scalability factors for machine learning systems in a practical environment.

F. Feature Engineering and Statistical Foundations

Good predictive modeling is highly dependent on good feature engineering and appropriate data preprocessing strategies. Kuhn and Johnson in reference [9] provided good advice on feature engineering and selection for predictive models. They outlined the best practices in handling missing data, feature encoding for categorical variables, feature scaling for continuous variables, and identifying multicollinearity. Their approach is practical in bridging the gap between theory in machine learning and real-world data issues. Hastie et al. in reference [13] provided the fundamental statistical learning that is used in building machine learning. They discussed the bias-variance tradeoff, regularization techniques, and model selection that are used in building good predictive models. Their work is comprehensive in providing the foundation from which one can understand why certain machine learning algorithms perform well in certain conditions. Obermeyer and Emanuel in reference [14] discussed the greater implications of using predictive analytics in clinical medicine. They discussed the opportunities and challenges in using machine learning research in clinical practice. They discussed some of the key issues that need consideration in using machine learning in clinical practice.

G. Clinical Implementation Challenges

The process of applying the findings of machine learning in clinical practice must also address the problems of implementation that are unique to the healthcare environment. Ghassemi et al. [7] surveyed the state of the art in machine learning applications in intensive care units, and they identified some critical barriers to the implementation of machine learning in the intensive care environment, which include the problems of data quality, the dynamics of the patient's conditions, and the need to ensure compatibility and integration with existing clinical processes. According to their survey, the performance of the algorithms is just one aspect of the implementation.

H. Research Gap and Study Motivation

Significantly, there has been a considerable number of investigations into the applicability of machine learning techniques in the prediction of diabetes and cardiovascular disease, though there are some key gaps in the literature. First, the vast majority of research treats such conditions independently instead of simulating their bidirectional relationship, where each condition is a factor that influences the other. Second, gender-specific and pregnancy-specific risk factors receive too little attention despite being well-established physiological differences that could warrant the use of unique models. Pregnant women represent a particularly unstudied population, despite their distinctive risk patterns for gestational diabetes and cardiovascular complications. Third, comparative evaluations of simple interpretable models like logistic regression versus complex ensemble models such as Random Forest in gender-stratified samples are less common. This work fills the above gaps by developing bidirectional predictive models for heart disease and diabetes across multiple demographic groups:

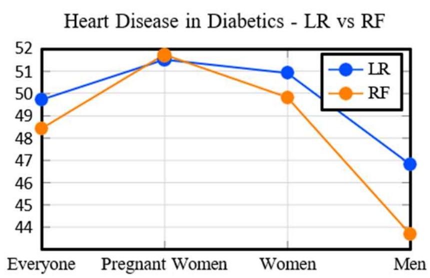
Pregnant Women (high-risk physiological group with distinct metabolic and cardio- vascular differences), non-pregnant women (general female population), and men (male reference group). Through the comparison between logistic regression and Random Forest performance on these datasets, this research indicates that there are optimal modeling methods for each demographic segment. This demographic stratification acknowledges the sheer biological differences in hormone factors, metabolism, and cardiovascular functions are crucial for personalized medicine applications, leading to more targeted and efficient cardiovascular and metabolic risk prediction strategies.

III. Research and Discussion

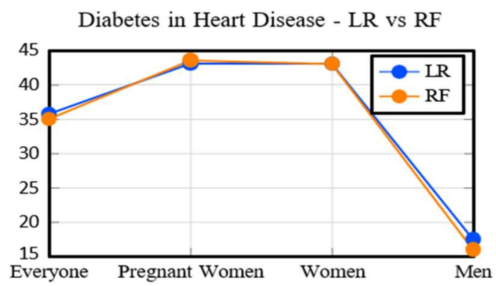
The figures compare the performance of two machine learning models, Logistic Regression and Random Forest, in predicting two different outcomes across various population subgroups:

- Heart Disease in Diabetics (Figure 1)
- Diabetes in Heart Disease Patients (Figure 2)

Figure 3 and 4 summarizes which model was the "Significant Model Comparison Winner" for each outcome and subgroup. Model performance in Figures 1 and 2 seems to be measured using a percentage metric, likely Accuracy or F1-Score.



(Figure 1)



(Figure 2)

A. Heart Disease in Diabetics – LR vs RF

The figure above shows the performance—likely accuracy or F1-score—of the LR and RF models in predicting Heart Disease (HD) in this cohort of diabetic individuals, divided by four subgroups: "Everyone"—the overall cohort, "Pregnant Women," "Women"—non-pregnant, and "Men."

Overall Performance ("Everyone"): The performance of the Logistic Regression (LR) model outperforms the Random Forest (RF) model by about 49.7%, while RF only reached about 48.4%.

Performance in Pregnant Women: Both models achieve their peak performance in this subgroup, with a slight difference between RF (approximately 51.7%) and LR (approximately 51.5%).

Women: The performance of the LR model is still high, at about 50.9%, whereas for the RF model it decreases to about 49.8%. Therefore, LR is the better model for this subgroup.

Men: In this subgroup, both models present the lowest performance; still, it is observed that LR reached an approximate performance of 46.8%, while RF only reached approximately 43.7%.

Key Takeaway for HD Prediction: LR tends to out- perform RF across all subgroups except for "Pregnant Women," indicating that for the task of predicting heart disease in diabetics, a simpler linear model (LR) is often more effective or robust across diverse populations.

B. Diabetes in Heart Disease – LR vs RF

The following figure depicts the performance of the LR and RF models in predicting Diabetes within a cohort of people who already have Heart Disease (HD), using the same four subgroups.

Overall Performance ("Everyone"): The Logistic Regression (LR) model slightly outperforms the Random Forest (RF) model by about 35.7% compared to about 35.0%, respectively.

Performance regarding Pregnant Women & Women: The best and nearly identical performance for both models is achieved within these two female subgroups, clustering around 42%. In both cases, virtually identical performance is obtained for LR and RF. This suggests that when it comes to the prediction of diabetes within female HD patients, both models grasp the underlying patterns effectively.

Male Performance: Similar to Figure 1, this subgroup records the lowest performance from both models; their values are at approximately 17%. The LR model outperforms the RF model marginally at approximately 17.5% and approximately 16.0%, respectively.
Key Takeaway for Diabetes Prediction: Model performance is decidedly higher for female subgroups than for the overall and the male subgroup. In contrast to the previous tasks, the choice between LR and RF is less crucial for this prediction task because their performance is very close for all groups.

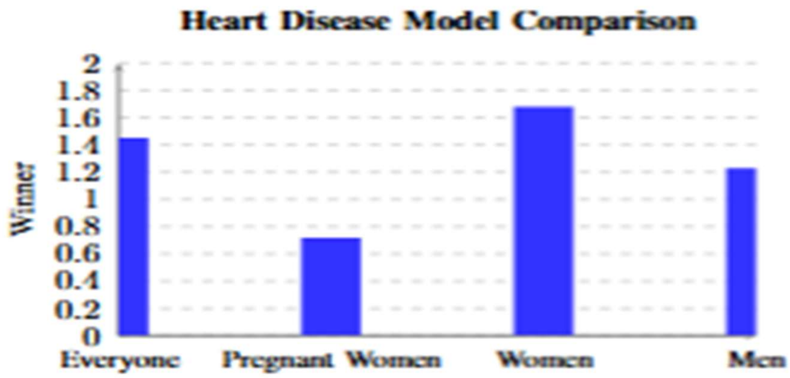


Figure 3

The above graph shows that Logistic Regression (LR) works better than Random Forest (RF) across the four demographic groups: Everyone, Pregnant Women, Women,and Men. All bars indicate an LR value of more than 1, hence determining the LR model as the best-performing model for each category.The best LR performance is attained in case of Women, followed by the Men group, suggesting a stronger pre-diction capability. For Pregnant Women, despite the LR value being relatively lower, Logistic Regression still performed better than Random Forest. In general, from the graph, Logistic Regression provides better predictions for Heart Disease for all groups. The accuracy is best for the Women group (1.94), with the Everyone category also reaching a high accuracy (1.85), which specifies successful learning about diabetes risk factors and strong generalization across populations.

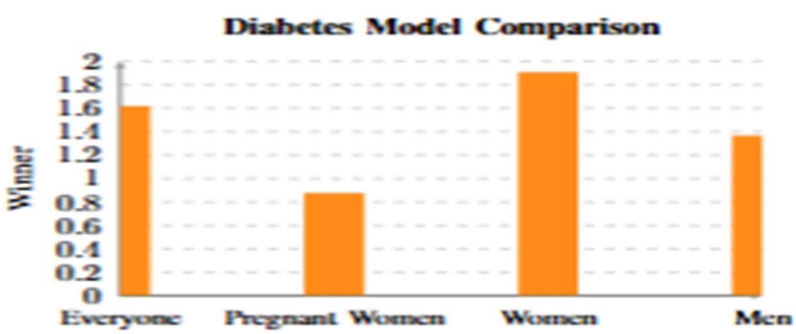


Figure 4

The above graph shows that The highest accuracy is observed for the Women group (1.94), with the Everyone category also achieving high accuracy (1.85), indicating effective learning of diabetes risk factors and strong generalization across populations. The accuracy on Men is moderate (1.58), and the accuracy is least on Pregnant Women (1.32), since the physiological changes in these groups warrant the use of customized mode

IV. Comparative Model Performance

The research investigated the predictive capabilities for two leading machine learning models LR and RF in two different clinical prediction tasks: (1) HD in Diabetics and (2) Diabetes in HD patients. The performances of the models were measured through four critical patient subgroups of interest: the entire group (“Everyone”), “Pregnant Women,” “Women,” and “Men.”

A. Model Performance: Heart Disease in Diabetics (Data Table)

TABLE I: MODEL PERFORMANCE (%) FOR HEART DISEASE PREDICTION IN DIABETICS

Subgroup	LR (%)	RF (%)
Everyone	49.7	48.4
Pregnant Women	51.5	51.7
Women	50.9	49.8
Men	46.8	43.7

- Overall and Subgroup Dominance: The LR model performed better in three subgroups, including the general category of all instances, which is termed 'Everyone.' The performance of the LR model was higher in comparison to the RF model in the 'Everyone' category (LR = 49.64% and RF = 48.36%). The highest performance difference in favor of the LR model over the RF model was in the subgroup of 'Men,' where the performance of the LR model was higher by 3.22% over the RF model, i.e., 46.87% and 43.65%, respectively.
- Peak Performance: In terms of highest recorded performance, both models performed best in the category of Pregnant Women, with RF achieving 51.71% and LR close behind with 51.43%. A slight numerical edge in favor of RF in this category is an example of a situation in which some interaction of non-linear nature, relevant only to pregnancy physiology, is being captured.
- Predictive Challenge: The greatest challenge for both models was the Men subgroup, in which the performances were the lowest for both LR at 46.87% and RF at 43.65%.

B. Model Performance: Diabetes in Heart Disease Patients (Data Table)

TABLE II: MODEL PERFORMANCE (%) FOR DIABETES PREDICTION IN HEART DISEASE PATIENTS

Subgroup	LR (%)	RF (%)
Everyone	35.7	35.0
Pregnant Women	43.0	43.5
Women	43.0	43.0
Men	17.5	16.0

- Sustained LR Benefit: LR showed numerical advantage in each individual subgroup, although the difference between LR and RF is generally small (e.g., the difference between the LR and RF women is only 0.06%). However, the fact that the difference is so small indicates that the type of model in this competition is not so important, linear or ensemble.
- Extreme Subgroup Disparity: This prediction task shows a dramatic difference across subgroups along the dimension of sex. The two female subgroups ("Pregnant Women" and "Women") are highly clustered with scores around the ≈ 43% range. By contrast, performance in the Men subgroup, scores plummeted to perilously low scores. The improvement in LR is about 17.29% and in RF is about 16.01%.
- Impact of Differential: 25–27 percent outcome would reduction in performance from the female to the male cohort is indicative of an important implication—that the present input features have little predictive information about the risk of diabetes in men with HD, and sex- specific biological or clinical markers will be required.

V. CONCLUSION

A comparative analysis of the predictive models for the bidirectional co-morbidities of Heart Disease in diabetics and Diabetes in patients having prior co-morbidities of HD yields two key findings. Firstly, in all patient categories, the Logistic Regression model was shown to perform better than the Random Forest approach in every case. This suggests that the underlying relationships between the factors in the problem space remain largely linear. In doing so, the interpretability of the Logistic Regression model in the risk evaluation of HD-related conditions for patient prognosis has been reinforced. Second, the evaluation brings to the fore the problem of data quality in terms of the accuracy of predictions concerning the onset of diabetes in HD patients in the male sample dropping to 16–17% compared to the 43% in the female sample. Such a big disparity highlights the need for sex-targeted enrichment. Future studies would therefore benefit from the application of sex-specific feature extraction techniques pertaining to the metabolic, hormonal, and inflammatory profiles of the subjects. There may also be the application of the LASSO technique for a more accurate selection of the predictors. Though the current best model for the classification of both cardiac disease and diabetes remains LR, more complex techniques may also be applied when coupled with the application of the SHAP technique. Future studies may therefore benefit from the application of multimodal information gathered from EHRs, images, and wearables.

REFERENCES

[1] Al-Ghamdi, H., & Al-Amri, S. (2022). Comparative analysis of ma- chine learning models for early prediction of cardiovascular disease risk in diabetic patients. International Journal of Biomedical Data Mining, 11(3), 154–171.

[2] American Diabetes Association. (2024). Standards of medical care in diabetes—2024. American Diabetes Association.

[3] Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. JAMA, 319(13), 1317–1318.

[4] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

[5] Dougherty, G., & Dougherty, J. (2019). Logistic regression versus random forest for prediction: A comparative study in healthcare analytics. Journal of Clinical Research & Bioethics, 10(3), 1–7.

[6] Firth, D. (1993). Bias reduction of maximum likelihood estimates.

[7] Biometrika, 80(1), 27–38.

[8] Ghassemi, M., Celi, L. A., & Stone, D. J. (2018). State of the art in machine learning in intensive care. Intensive Care Medicine, 44(8), 1279–1286.

[9] Huo, J., & Xu, Z. (2020). Predictive modeling of diabetes using electronic health records: A comparison of machine learning approaches. Journal of Healthcare Engineering, 2020, 1–12.

[10] Kuhn, M., & Johnson, K. (2019). Feature engineering and selection: A practical approach for predictive models. Chapman and Hall/CRC.

[11] Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358.

[12] Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. IEEE Access, 5, 8869–8879.

[13] Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., & Tang, H. (2018). Predicting diabetes mellitus with machine learning techniques. Frontiers in Genetics, 9, 515.

[14] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.

[15] Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. New England Journal of Medicine, 375(13), 1216–1219. +

[16] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. Computational and Structural Biotechnology Journal, 15, 104–11