Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

To Develop and Validate An AI Driven Logistic Regression Model in Patients with Metabolic Risk Factors using ECG Parameters

Ms. Zoya Zabeen 1st
Department of Medical Physiology, Amity Institute of Health Allied Sciences, Amity University, Noida, Uttar Pradesh, India

Dr. Satyajit Tripathy 2nd
Assistant Professor, Department of Physiology and Allied Sciences, Amity University, Noida, Uttar Pradesh, India

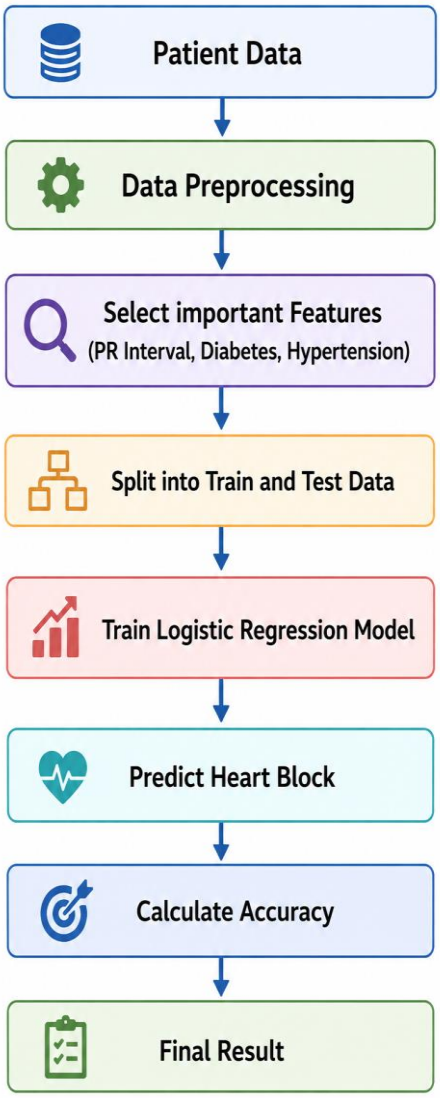
Abstract **- Early identification of cardiac conduction abnormalities in metabolic diseases may improve the effectiveness of the treatment. The objective of this work is to develop a model for the prediction of cardiac block in patients having some of the risk factors such as diabetes mellitus, hypertension, and PR interval observed on the electrocardiogram readings. Logistic regression analysis was performed to identify the factors contributing to the occurrence of cardiac block**

Keywords - Heart Block, Logistic Regression, PR Interval, Type 2 Diabetes Mellitus, Predictive Modeling, Clinical Telemetry

I. INTRODUCTION

Since cardiac conduction disturbances, like heart blocks, occur when electrical impulses between the atria and ventricles are delayed or disrupted, it is critical to identify them early. Traditionally, idiopathic degenerative changes of the specialized conduction system were observed in older populations. Changes in lifestyle have led to the early remodeling of the conduction system of the heart. Particularly, chronic hypertension and Type 2 Diabetes Mellitus have now been added to the list of the causes of structural and electrophysical changes in the heart. This complex cardiovasculopathy is uniquely dangerous because it often progresses silently with few warning signs before developing into a high-grade atrioventricular block or symptomatic bradyarrhythmias. Changes in conduction are still most commonly evaluated through the analysis of a 12 Lead Electrocardiogram. The PR Interval is one of the most critical components of an Electrocardiogram.[1] It measures the time it takes for all of the electrical activity to pass through from the Atria to the Ventricles via the AV node and the His-Purkinje System. Traditionally, during the evaluation of a 12 Lead Electrocardiogram, a first degree heart block is diagnosed when the PR Interval is above a set threshold. However, in patients with metabolic syndrome, small changes in the length of the PR Interval, even within the threshold, can be highly predictive. Developments in health informatics have led to increasing interest in the use of machine learning algorithms for automatic screening and triage. Unfortunately, deep learning networks, despite their classification accuracy, are complex black box systems which hide from the clinician the physiological interrelations. In contrast, binary logistic regression offers an interpretable, transparent, and straightforward system. With this method, clinical researchers are able to retrieve untransformed model coefficients and literally convert them to clinically relevant odds ratios. In this paper, we analyze how we designed a transparent predictive model for assessment of the risk of heart block and, then, weighted it and analyzed it physiologically.[2]

II. WORKFLOW



(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

Fig. 1. Block Diagram of data visualisation

III. PROPOSED ALGORITHM

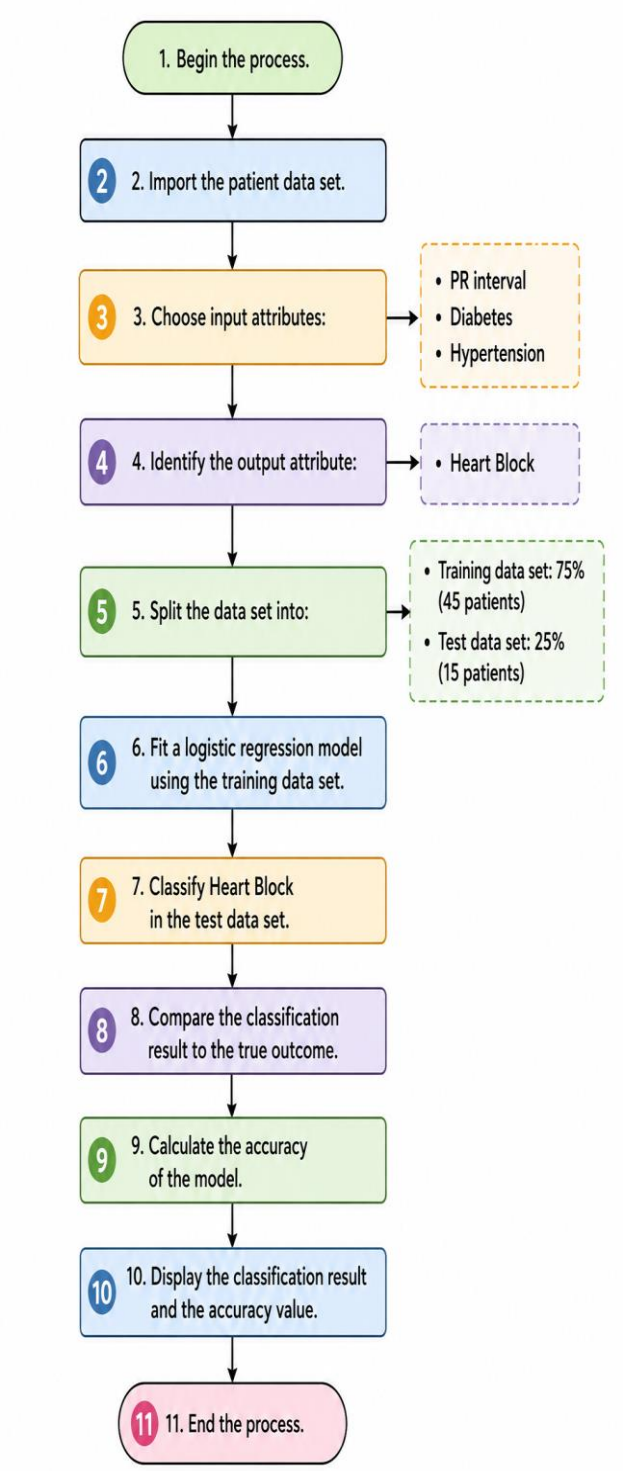


Fig. 2. Proposed Classification Process

IV. MATERIAL METHOD

The data used in this research were collected from the Cardiology Department of Medicare Hospital, Patna, after obtaining the approval of the hospital management. The initial set of data included telemetry data of 6

TABLE I. PATIENT DATA

Patient Dataset	Patient Details			
	PR Interval	Diabetes	Hypertension	Heart Block
1	155	1	0	0
2	158	0	1	0
3	161	1	0	0
4	163	1	1	1
5	166	1	0	0
6	168	0	1	0
7	171	1	0	1
8	173	0	1	1
9	175	0	0	0
10	178	0	1	0
11	180	1	0	1
12	182	0	0	0
13	185	1	1	1
14	187	1	1	0
15	190	1	0	0
16	192	0	0	0
17	194	1	1	1
18	196	0	0	0
19	198	1	1	1
20	200	0	1	1
21	170	0	1	1
22	172	1	1	1
23	176	0	1	0
24	179	1	1	1
25	181	0	1	0
26	184	1	1	1
27	186	0	1	0
28	188	1	0	1
29	191	0	1	1
30	193	1	1	1
31	195	1	1	1
32	197	1	0	0
33	199	0	1	1
34	202	1	1	1
35	204	0	1	0
36	206	1	1	1
37	208	0	1	0
38	210	1	1	1
39	212	0	1	1
40	215	1	1	1
41	201	1	1	1
42	203	0	1	1
43	205	1	1	1
44	207	0	1	1
45	209	1	1	1
46	211	0	1	1
47	213	1	1	1
48	216	0	1	1
49	218	1	1	1
50	220	0	1	1
51	222	1	1	1
52	225	0	1	1
53	227	1	1	0
54	229	0	1	1
55	231	1	1	1
56	233	0	1	1
57	235	1	1	1
58	237	0	1	0
59	239	1	1	1
60	242	0	1	1

V. RESULT

The logistic regression model accurately predicted the result for 80% of the patients in the test set. The ROC-AUC was 0.84, indicating good separation of patients with normal conduction and those with heart block.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

The model coefficients and corresponding odds ratios were

TABLE II. LOGISTIC REGRESSION MODEL

Predictor variable	Raw model coefficient	Exponential odds ratio	Clinical interpretation
PR interval duration (per 10-ms increase)	0.285	1.33	Incremental increase in conduction risk
Type 2 diabetes mellitus	1.163	3.20	Primary metabolic risk multiplier
Systemic hypertension	0.531	1.70	Secondary structural strain factor
Model constant (intercept)	−3.421	Not applicable	Baseline log-odds reference value

.

With every increase in the PR interval duration by 10 milliseconds, the odds of heart block were expected to increase by 33%, assuming that the other variables in the model remain constant also statistical analysis revealed that the PR interval duration at baseline and the disease of type 2 diabetes mellitus were the most important predictors in the model. Patients with diabetes had 3.2 times more chances of having heart block as compared with those without diabetes, Systemic hypertension conveyed an odds ratio of 1.70, which translates into a 70% higher odds of having heart block and type 2 diabetes mellitus revealed the highest odds ratio from among all the comorbid conditions included in the model. Overall, Type 2 diabetes mellitus and the PR interval duration emerged as the most important predictors in the fitted model. As compared with that of hypertension the odds ratio for diabetes was much higher

VI. MODEL EVALUATION REPORT

TABLE II. Model Evaluation Results on the Test Dataset

Patient	PR Interval	Diabetes	Hypertension	Actual Heart Block	Predicted Heart Block	Result
7	171	1	0	0	0	Correct
9	175	1	0	0	0	Correct
14	187	0	1	0	0	Correct
23	176	0	1	0	0	Correct
29	191	0	1	1	0	Wrong
33	199	0	1	1	1	Correct
34	202	1	1	1	1	Correct
37	208	0	1	0	1	Wrong
42	203	0	1	1	1	Correct
45	209	1	1	1	1	Correct
47	213	1	1	1	1	Correct
48	216	0	1	1	1	Correct
50	220	0	1	1	1	Correct
51	222	1	1	1	1	Correct
58	237	0	1	0	1	Wrong

Score: 12 correct out of 15 = 80% accuracy

For the current assessment, a test set of 15 patients was created as a subset of the original sample. Test instances, as seen below, were not seen by the algorithm during its learning. As it follows from the above-mentioned table, for 12 out of 15 test patients the prediction of "Predicted Heart Block" correlated with the true result, which provided the overall accuracy of 80%. The following three mistakes occurred:

for patient 29 the model predicted "no" (heart block was not predicted) while in reality, there was a heart block (false negative);

for patients 37 and 58 the model predicted "yes" (heart block) despite the lack of conduction abnormality (false positive).

As can be noticed, the model predicts heart block when PR interval is high and hypertension is observed. It correlates with the general pattern observed in most cases, though it does not cover all of them. Thus, we have the model with the slight over-prediction of heart block. Overall, this is the preliminary analysis conducted on a rather small dataset of 60 patients from one hospital. Despite the fact that results seem to be promising, the model has to be externally validated on a much larger sample of patients.

VII. DISCUSSION

The performance metrics of this model point out an important clinical consideration. It is possible to gain significant predictive power by combining simple and common variables without the need for patients to undergo costly and complex invasive electrophysiological tests. The final Area Under the Curve score of 0.84 indicates that the model and its components easily fall within the excellent discrimination range, making the model viable for early screening in out patient settings.

The result of the odds ratios and the physiological associations show that the progressive nature of the delay of the atrioventricular node can be represented in terms of atrioventricular nodal delay of a few milliseconds. The delay is not a sharp all-or-nothing phenomenon. It can be viewed as a reflection of gradual microstructural changes, strain or subtle inflammation occurring in the atrioventricular node. It is telling that the presence of type 2 diabetes mellitus amplifies the risk beyond that of systemic arterial hypertension. It seems that chronic hyperglycemia is more damaging to myocardial tissue and the specialized conduction system tissue than chronic arterial hypertension. The latter causes strain of the atrial walls and hypertrophy that in time alters the normal pathways for

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

cardiac conduction.[3] In contrast, type 2 diabetes mellitus is an amalgam of various pathologies. In particular, the presence of advanced glycation end-products in the myocardial extracellular matrix, hyperglycemia and the resultant interstitial fibrosis and oxidative stress and so micro-remodeling that also impacts the bundle branches and the Purkinje network may explain the high odds ratio in this study. Machine Learning models and systems have focused on formulating data parameters and translating them into Odds Ratios (ORs) to draw patterns to aid health care's physiological processes. The ORs describe the risk and help the clinicians decide the telemetric monitoring needs, etc., to draw the assistance on the risk and help regulate the monitoring protocol, etc. for the high risk patients.

VIII. ETHICAL FRAMEWORK AND CLINICAL DATA COLLECTION

The study was conducted over a structured seven-week clinical training period in the non-invasive telemetry unit of the Department of Cardiology at Medicare Hospital, Patna. Clinical observations, patient-record reviews, and diagnostic telemetry assessments were carried out within the unit. To develop the prediction model, we reviewed the clinical records of adult patients with predefined metabolic risk factors. In order to get the relevant clinical and ECG data about heart block, we filtered the data.[3][4]

IX. OUTCOME AND PREDICTOR VARIABLES

Well-defined results were obtained. A clear yes/no variable was created for the model to show whether the patient had heart block according to the clinical data.

Three predictors were used in the calculation:

[1] The length of the PR interval measured in milliseconds was taken from the ECG.
[2] It was established if the patient had type 2 diabetes or not.
[3] The presence of hypertension was observed as a binary variable depending on whether the diagnosis of this disease was active in the patient.

X. MODEL DEVELOPMENT AND IMPLEMENTATION

In our study, we have chosen binary logistic regression as our method of analysis. This method accepts many independent variables as input and provides the result from 0 to 1, that is probability of the getting the result. Using the data preprocessing methods in Python we first cleaned the dataset and removed outliers to create two independent sets, training set and validation set, so that the model could be tested on entirely new data.[5]

X. COEFFICIENT EXTRACTION AND CALCULATION OF ODDS RATIOS

If the value is greater than 1 means there are greater odds for a heart block. For any value less than 1, the odds decline. This process of changing the values into risk multipliers, allowing both physicians and patients to interpret the results easier. After the model has been fitted, data from Python is used here. One should note that although logistic regression provides the coefficients in terms of odds in a logarithmic manner (that makes it difficult for possible applications), all coefficients are exponentiated using NumPy, thus gaining the probability ratios. In this case, it is easy to see whether the coefficient is greater or less than 1. If it is greater than 1, this factor increases the odds for a heart block, and if the number is smaller than 1, the risk is decreased.[6][7]

XI. MODEL VALIDATION

The choice of the ROC-AUC as a metric for computing the performance of the model in classifying patients with and without heart block was made because it does not depend on the threshold which is set. To evaluate the performance of the model, we applied it to an independent test set. We evaluated the number of patients that the model correctly classified (i.e., its accuracy).[8]

XII. CONCLUSION AND FUTURE WORK

In the study, an interpretable logistic regression model was developed to estimate the probability of heart block in adults with metabolic risk factors. Using three easily obtainable parameters—PR interval length, type 2 diabetes condition, and systemic hypertension—validation accuracy was 80.0% and ROC-AUC metric value was 0.84. The analysis shows that long PR intervals and type 2 diabetes are especially predictive of conduction abnormalities and odds ratio transformation of coefficients improves the interpretability of the results. In conclusion, easily interpretable machine-learning algorithms based on clinical data could assist in the prediction of the risk of heart block. However, the current study was conducted on the relatively small dataset of 60 patients and external validation on a larger patient sample is needed before any routine application of the model in the clinical practice. Future studies should focus on obtaining larger dataset and testing the performance of other algorithms like Random Forest and Support Vector Machine. In addition, the inclusion of sex, QRS and QT intervals and medications could improve accuracy of the model.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 08 , August - 2026

REFERANCES

[1] Kavya S M, PrathanyaSree C, Deepasindhu M, Nowshika B, Shijitha R, Heart Disease Prediction Using Logistic Regression, JCLMM 1/11 (2023) |573–579.

[2] Ambrish G, Bharathi Ganesh, Anitha Ganesh, Chetana Srinivas, Dhanraj, Kiran Mensinkal,Logistic regression technique for prediction of cardiovascular disease, Global Transitions Proceedings Volume 3, Issue 1, June 2022, Pages 127-130.

[3] Symeon Savvopoulos 1, Haralampos Hatzikirou 1,2, Herbert F Jelinek, AI-based integration of ECG biomarkers for assessing cardiac risk in type 2 diabetes mellitus with comorbid conditions for patient stratification, Volume 12, Front. Med., 08 September 2025.

[4] Ambrish G,Bharathi Ganesh, Anitha Ganesh, Chetana Srinivas, Logistic Regression Technique for Prediction of Cardiovascular Disease, April 2022Global Transitions Proceedings 3(1).

[5] Yagyanath Rimal1, Navneet Sharma, Siddhartha Paudel, AbeerAlsadoon,Madhav Parsad Koirala & SumeetGill,Comparative analysis of heart disease prediction using logistic regression, SVM, KNN, and random forest with cross-validation for improved accuracy,Scientific Reports volume 15, Article number: 13444 (2025).

[6] Alina Ioana Forray, Mădălina Adina Coman, Ruxandra Simonescu-Colan, Andreea Isabela Mazga, Răzvan Mircea Cherecheș, Cristina Maria Borzan. The global burden of type 2 diabetes attributable to dietary risks: insights from the global burden of disease study 2019. Nutrients. (2023) 15:4613. doi: 10.3390/nu15214613.

[7] Namazi N,Moghaddam SS, Esmaeili S, Peimani M, Tehrani YS, Bandarian Fet al., Burden of type 2 diabetes mellitus and its risk factors in North Africa and the Middle East, 1990–2019: findings from the global burden of disease study 2019. BMC Public Health. (2024) 24:98. doi: 10.1186/s12889-023-16540-8

[8] F.A. Latifah, I. Slamet, Comparison of heart disease classification with logistic regression algorithm and random forest algorithm Proceedings of the AIP Conference (2020), p. 2296

(This work is licensed under a Creative Commons Attribution 4.0 International License.)