Electronic Journal of Education, Social Economics and Technology
Vol. 7, No. 1, (2026), e1048
ISSN 2723-6250 (online)
DOI: https://doi.org/10.33122/ejeset.v7i1.1048

SAINTIS publishing

Research Article

Application of the Decision Tree Algorithm for Early Detection of Heart Disease Based on IoT

Rosa Englina Silaban*, Ridho Maulana Siregar, Natasya Aulia Angkat, Mhd. Raihan M. Manurung, Achmad Ridwan

Universitas Prima Indonesia, Medan, Indonesia, 20153
*Corresponding Author: rosasilaban02@gmail.com| Phone: 085381877252

ABSTRAK

Heart disease is one of the leading causes of death worldwide, accounting for 32% of all global deaths. Technological developments, particularly in the Internet of Things (IoT), enable real-time monitoring of heart health and early warning alerts. This study aims to implement a Decision Tree algorithm to classify patient conditions based on vital parameters, including BPM, SpO_2, systolic and diastolic blood pressure, and body temperature. The model was trained using a vital parameter dataset and evaluated using a confusion matrix, ROC curve, and feature importance. Test results show that the Decision Tree model achieves an accuracy of 85% with a macro-AUC value of 0.448. These results prove that the Decision Tree algorithm can be used for patient condition classification with reasonably good performance, although the model still tends to make prediction errors in some minority classes.

Keywords: Heart Disease; Internet of Things; Decision Tree

1. INTRODUCTION

The heart is responsible for pumping blood that carries oxygen and nutrients to all cells in the body for survival (Ryfai, Hidayat, and Santoso 2022). Non-communicable diseases (NCDs) now account for a significant proportion of global mortality. One of the NCDs with the most significant contribution is heart disease, including coronary heart disease, cardiovascular disease, and other types. Globally, cardiovascular disease accounts for approximately 32% of all deaths (Mensah, Fuster, and Murray 2024). In Indonesia, the results of the 2014 Sample Registration System (SRS) survey show that coronary heart disease is the leading cause of death in all age groups, with a proportion of 12.9% (Ryfai et al., 2022). Therefore, technology that can monitor heart health early on is needed. One technology that supports this is the Internet of Things (IoT). The Internet of Things is a technology that enables control, communication, and collaboration with various hardware devices (Ayu Syahfitri, 2025). By utilizing ESP 32 as a microcontroller that has been integrated with a Wi-Fi module, this system can be accessed and controlled remotely in an efficient manner (Mikrokontroler Esp 32 Sebagai Alat Monitoring Pintu Berbasis Web 2022). In addition to the ESP32 microcontroller, this study utilizes several sensors to support the monitoring process, including MAX30102, GY-906 (MLX90614), and AD8232. The MAX30102 sensor is used to monitor heart rate (beats per minute / BPM) and blood oxygen levels (SpO_2), while the GY-906 (MLX90614) is used to measure body temperature. Meanwhile, the AD8232 displays a simple heart wave, providing an overview of the heart's electrical activity. Therefore, this study uses a separate device to measure blood pressure so that the data obtained can complement the monitoring results from the sensors used.

This study aims to develop an IoT-based early detection system for heart disease using the Decision Tree algorithm as a health data classification model. The benefits of this research include improving the quality of health services with the system and accurate early detection. This study is expected to improve the quality of health services so that medical personnel can make more accurate and faster decisions. It also aims to increase public awareness about heart health so that people are more aware of the risks of heart disease and the importance of preventive measures. Contribution to research in the field of health and technology, particularly in the application of IoT and machine learning for early disease detection, as well as a reference for research in the same field (Karamina & Ilham, 2025). Increased accessibility to health monitoring allows more people to access quality health services without visiting health facilities directly. This research is limited to monitoring vital parameters such as heart rate (BPM), blood pressure, body temperature, and SpO2 (oxygen saturation).

The novelty of this research lies in the combination of IoT and Decision Tree algorithms in a real-time monitoring platform that can be used directly by users and medical personnel. Several previous studies have used algorithms such as

Random Forest (92.63%) (Selayanti et al., 2024), Linear Discriminant (81.22%) (Rashad, Isnanto, and Widodo, 2023), and Support Vector Machine (85%) (Hidayat et al., 2024). And Rough Neural Network (88.52%) (Dataset and Network 2024).

2. RESEARCH METHOD

This research method uses Research and Development (R&D) in to develop early detection of heart disease. The **Figure 1** shows a sequence of activities that can facilitate research. The stages involved include problem identification, planning, product development, product testing, and revision evaluation.

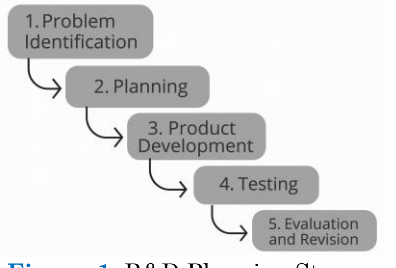

Figure 1. R&D Planning Stage

Problem identification was carried out through literature studies and health observations to understand the problems patients face in detecting heart disease early. The main problem found was inefficiency, and manual responses to heart disease did not worsen the condition of the heart. Based on the needs analysis results, the tool was designed to receive information in real time.

At this stage, the device was designed using Fritzing to produce a 2D model. The main components designed include the ESP32, MAX30102 sensor, GY906 sensor, AD8232 sensor, 20x4 I2C LCD, and others.

Table 1. Tools and Materials

No.	Tools/Materials	Function
1	ESP32	For long-distance communication and data transfer
2	AD8232	Signal conditioning module for applications
3	MAX30102	To measure heart rate and blood oxygen levels (SpO2)
4	LCD 20 x 4 I2C	Alphanumeric LCD display with 20 x 4 characters
5	Electrode Pads	Deliver electrical impulses to the body
6	Jumper Wires	Connect components in the circuit
7	USB	Connects the device to other devices
8	GY 906	For non-contact temperature measurement
9	On/Off Button	Functions as a switch to connect or disconnect electric current in a circuit
10	Power Supply	Ensures the device receives a stable and safe power supply as needed, either in DC or AC form

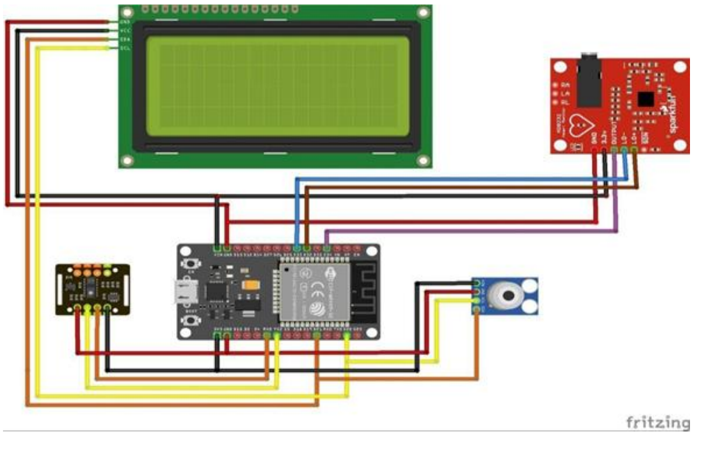

Figure 2. Product Design

The image above shows the overall 2D design of the early detection device for heart disease. This system comprises various interconnected components that work together to achieve this goal. In designing the device, we calculated the accuracy of each sensor and the MAE to ensure the accuracy of the sensors used. Product development involved assembling mechanical, control, and electrical systems. The system includes manufacturing the device frame, installing the sensors, and using them. The control system was designed using IoT technology for real-time monitoring and ESP 32 integration, which allows for the monitoring of heart health.

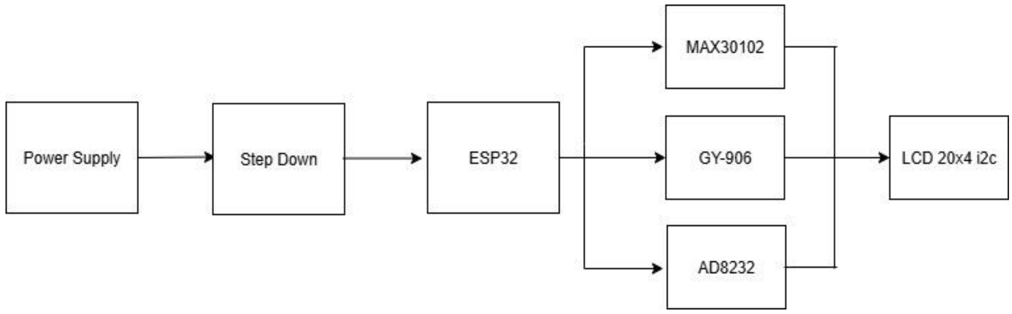

Figure 3. Hardware Block Diagram

Initial testing was conducted on each component separately to ensure performance before assembly and to calculate each sensor's accuracy and MAE. The early detection device for heart disease was tested to evaluate the device's performance in monitoring heart health. Testing was conducted while the body was at rest, for approximately 3–5 minutes. Before testing began, blood pressure was measured using a digital sphygmomanometer. The data generated was displayed on the Blynk platform during testing, and a preliminary analysis was presented using the Decision Tree algorithm. All test data was also automatically saved via Google Spreadsheet for further recording and analysis.

Table 2. Vital Parameters

Age	Gender	Systolic (mmHg)	Diastolic (mmHg)	BPM (beats/min)	SpO_2 (%)	Body Temperature (°C)
0–1 year	M & F	75–100	50–65	100–160	95–100	36.5–37.5
1–5 years	M & F	80–110	50–80	80–120	95–100	36.5–37.5
6–12 years	M & F	90–115	55–80	70–110	95–100	36.5–37.5
13–17 years	M & F	110–120	65–85	60–100	95–100	36.5–37.5
18–40 years	Male	110–130	70–85	60–100	95–100	36.5–37.5
	Female	105–125	65–80	60–100	95–100	36.5–37.5
41–60 years	Male	115–135	75–90	60–100	95–100	36.3–37.3
	Female	110–130	70–85	60–100	95–100	36.3–37.3
>60 years	Male	120–140	80–90	60–100	95–100	36.0–37.2
	Female	115–135	75–85	60–100	95–100	36.0–37.2

The vital parameter table displays the measurement results of each health indicator obtained from the sensor during the trial process. Each column represents the type of parameter measured, such as heart rate (BPM), blood oxygen level (SpO_2), body temperature, and systolic and diastolic blood pressure.

A decision tree is a data structure consisting of nodes and edges. Nodes in a tree are divided into three types: root nodes, branch/internal nodes, and leaf nodes (Nasrullah 2021). Accuracy in a decision tree is a measure that shows how often the decision tree model makes correct predictions compared to the total number of predictions made. Entropy is a quantitative measure used to describe a data set's uncertainty or irregularity level. Information Gain (IG) is a measure used in the Decision Tree algorithm to determine how much an attribute can reduce uncertainty (entropy) in the data when used to separate the data set. Precision shows how close the difference in values is when repeated. Recall is the ratio of optimistic correct predictions to the total positive correct data. F1-score means the weighted average comparison of precision and recall. Specificity can be interpreted as the accuracy of predicting negative compared to the total negative data. Error rate compares the number of incorrect data units to the total number of data units.

Fritzing is CAD-based open-source software for designing electronic hardware for beginners and hobbyists. Included in the Electronic Design Automation (EDA) category with a low level of difficulty, Fritzing facilitates the creation of schematics, PCB layouts, and digital project documentation. Google Colaboratory (Google Colab) is a cloud service that runs Python code interactively through a browser with free GPU and TPU support. This study used it to implement and train the Decision Tree algorithm and evaluate the model using a confusion matrix, ROC Curve, and feature importance analysis. Arduino IDE is the official software for writing, compiling, and uploading code to microcontrollers such as Arduino and ESP32. This study used it to program the ESP32 to read sensor data, send it to Blynk, and integrate control and measurement functions. Blynk is an IoT platform for controlling hardware, collecting sensor data, and monitoring results via an app or the web. This study used it to display real-time sensor data, provide preliminary analysis with a Decision Tree, and store data in Google Spreadsheets.

3. RESULT AND DISCUSSION

The MAX30102 is an optical sensor specifically designed for noninvasive measurement of heart rate and blood oxygen levels (SpO_2). This sensor works using the principle of photoplethysmography (PPG), a method of measuring changes in blood volume based on variations in light absorption by body tissue.

Table 3. MAX30102 Testing

No.	BPM MAX30102	BPM Pulse Sensor	SpO_2 MAX30102	SpO_2 Pulse Oximeter
1.	83	80	88	98
2	96	97	89	98
3.	103	113	89	97
4.	99	96	88	96
5.	93	85	90	98
6.	85	80	88	86
7.	93	94	91	95
8.	96	93	90	98
9.	94	95	88	98
10.	87	86	91	99

Based on the calculation results, the MAX30102 sensor has an accuracy rate of 96.2% for heart rate (BPM) measurements and 91.25% for blood oxygen level (SpO_2) measurements, with a Mean Absolute Error (MAE) value of 3.6 BPM and 7.5%, respectively. These results show that the MAX30102 sensor performs well and is reliable for real-time monitoring of vital parameters.

The GY-906 (MLX90614) is a non-contact infrared sensor (non-contact infrared thermometer sensor) used to measure the surface temperature of objects or bodies without the need for direct physical contact. This sensor uses the principle of infrared radiation, whereby every object with a temperature above absolute zero emits infrared energy, which is then detected by the thermopile element inside the sensor.

Table 4. GY906 Testing

No.	Suhu GY-906 (^{0}C)	Suhu Termometer (^{0}C)
1.	35,1	36,1
2	36	36,3
3.	36,4	36,7
4.	36,5	37,0
5.	36,1	36,6
6.	36	36,6
7.	35,4	36,5
8.	35,3	35,5
9.	36,1	36,9
10.	35,3	36,3

Based on the calculation results, the GY-906 (MLX90614) sensor has an accuracy rate of 98.28% and a Mean Absolute Error (MAE) of 0.62°C. These results show that the sensor can provide precise and stable temperature readings, especially after undergoing a heating process for ±3 minutes, making it suitable for monitoring body temperature in IoT-based healthcare applications.

Based on the evaluation results, the model produced an overall accuracy rate of 85%, which means that 23 out of 27 test data were correctly predicted. The macro average values for precision were 0.40, recall 0.55, and F1-score 0.44. Macro average is used to calculate the average metric for each class without considering the proportion of data in each class, thus providing a balanced overview of the model's performance across all classes.

The weighted average values for precision are 0.85, recall 0.85, and F1-score 0.84. The weighted average considers the amount of data in each class, thus providing a model performance assessment that better represents the actual distribution of test data. The calculation results also show an Entropy value of 1.7504 and Information Gain for the SpO_2 attribute of 1.1988, which are used to select the best attributes in the decision tree formation process.

	precision	recall	f1-score
(Adult male abnormal)		0.33	0.00
(Middle-aged abnormal)		0.00	0.00
(spO2 below normal)		1.00	1.00
Body temperature abnormal		1.00	0.83
.idlle-age female abnormal)		0.50	0.67
Not at Risk		0.00	0.00
accuracy		0.5	27
macro avg		0.85	0.44
weighted avg		0.85	0.84

Figure 4. Test Results

Based on the confusion matrix results, the Decision Tree model can classify several classes very well, but certain classes still have weaknesses. The Risk Class (SpO_2 below normal) performed flawlessly with 15 test data classified correctly without any prediction errors. The risk class (abnormal body temperature) also performed well, with six correct data points and only one prediction error, which was classified as risk (abnormal middle-aged women). Meanwhile, the Risky class (abnormal adult male) and Risky class (abnormal adult female) each had one correct prediction, but for the abnormal adult male class, there were still prediction errors for other classes, including Non-Risky. The lowest performance was found in the Risky (abnormal middle-aged men) and Not Risky classes, where all data in these two classes were incorrectly predicted to be in other classes. These results indicate that the model tends to work very well on classes with a large amount of data (majority classes), but has difficulty classifying classes with a small amount of data (minority classes), which is most likely due to the imbalance in the distribution of data in the dataset.

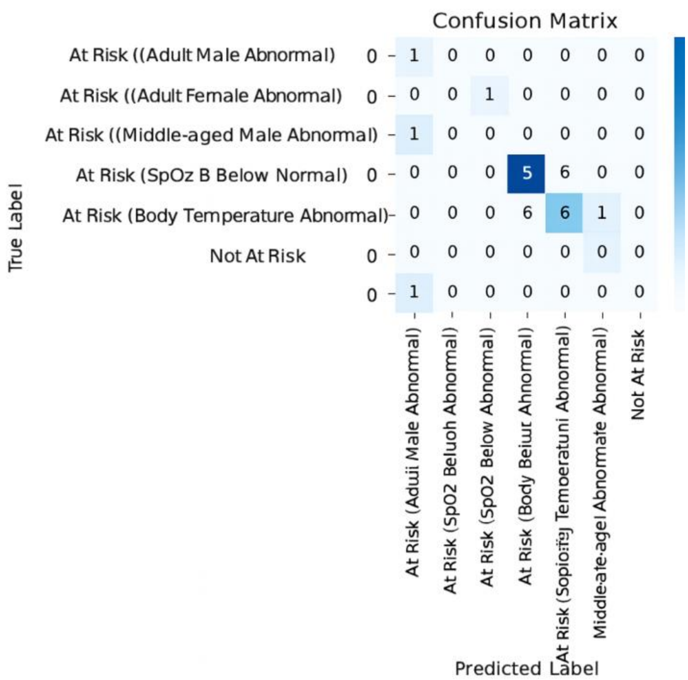

Figure 5. Confusuion Matrix

The learning curve Decision Tree shows that training accuracy is consistently high, ranging from 0.88 to 0.96, with a slight downward fluctuation around 60 data points before rising again to around 0.92–0.93 at the end of testing. The initial validation accuracy was low at around 10 data points, at 0.38, then increased rapidly at 20 to 40 data points, reaching a range of 0.82–0.85. The peak validation accuracy occurred at around 60 data points, with a value of around 0.86–0.87, and the smallest gap with the training accuracy. After that, the validation accuracy decreased in the range of 70 to 85 data points, falling to around 0.76–0.81, before finally rising partially back to the range of 0.82–0.83 at 95 to 105 data points. This pattern shows that the accuracy gap between training and validation is enormous at the beginning, narrows around the peak point, and widens again after the data count exceeds 70, indicating optimal model adjustment at that peak point.

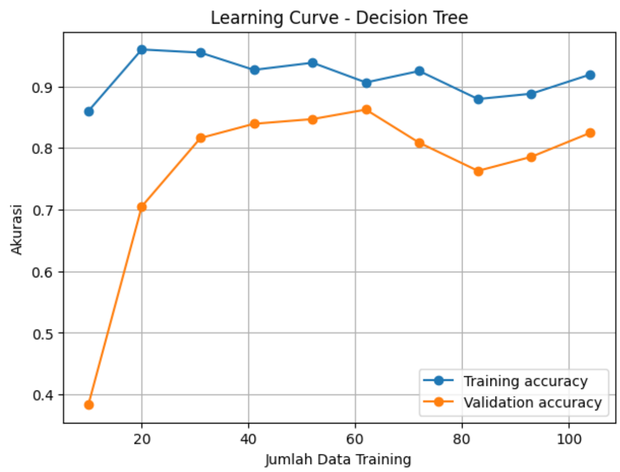

Figure 6. Learning Curve

The multiclass ROC curve in the image shows a macro-AUC value of 0.448, indicating that the model's overall performance is below the level of perfect random prediction (AUC=0.5). Some classes, such as *Abnormal adult women* and *Abnormal middle-aged women*, have an AUC of 0.500, equivalent to random prediction, while *Abnormal body temperature* has the lowest performance with an AUC of 0.250. Classes with relatively better performance are *Abnormal Adult Male* (AUC=0.481) and *Abnormal Middle-aged Male* (AUC=0.462), while *SpO_2 Below Normal* has an AUC of 0.442. The distribution of the curve, which tends to be close to the random line (dashed line), indicates that the model's ability to distinguish classes in most categories is still low.

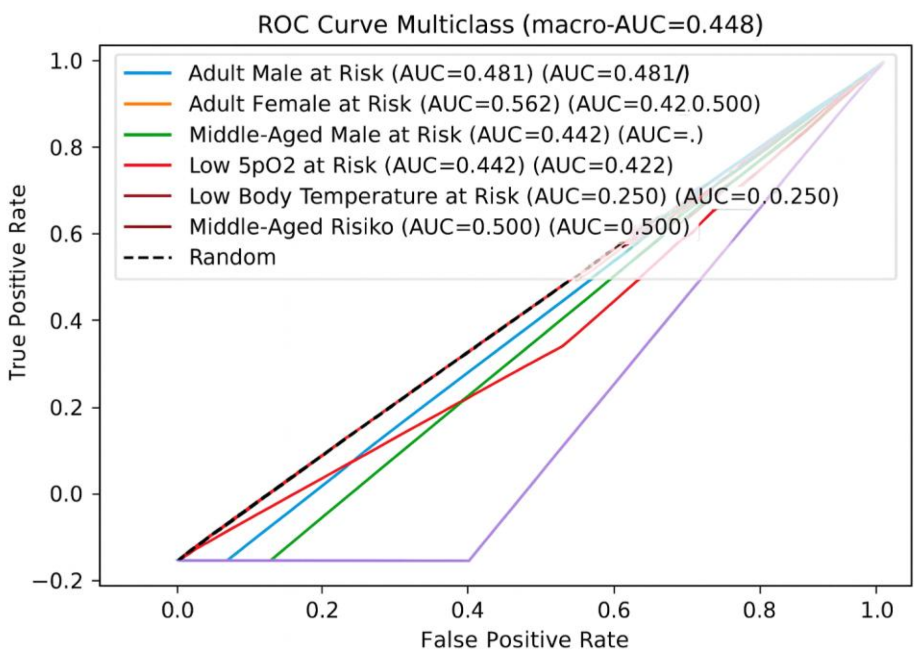

Figure 7. ROC Curve

The decision tree in the figure below shows the flow of patient condition classification based on vital parameters, with SpO_2 as the first separating variable. The initial node separates the data based on the threshold $SpO_2 \leq 94.5$. If SpO_2 is below or equal to this value, the following classification is based on a Systolic value ≤ 112. This branch has a Gini value of 0.0 at both end branches, which means that the entire sample in that node is in the same class ("At Risk").

If $SpO_2 > 94.5$, the process continues with checking BPM ≤ 155. For data with BPM lower than or equal to this threshold, separation is based on body temperature ≤ 36.45. Data with a temperature ≤ 36.45 is mainly classified as "At Risk" with a Gini value close to zero, indicating high data purity. If the temperature is higher, the following classification is performed on the Systolic parameter ≤ 141.5, then Diastolic ≤ 60, before reaching the leaf node with a low Gini value.

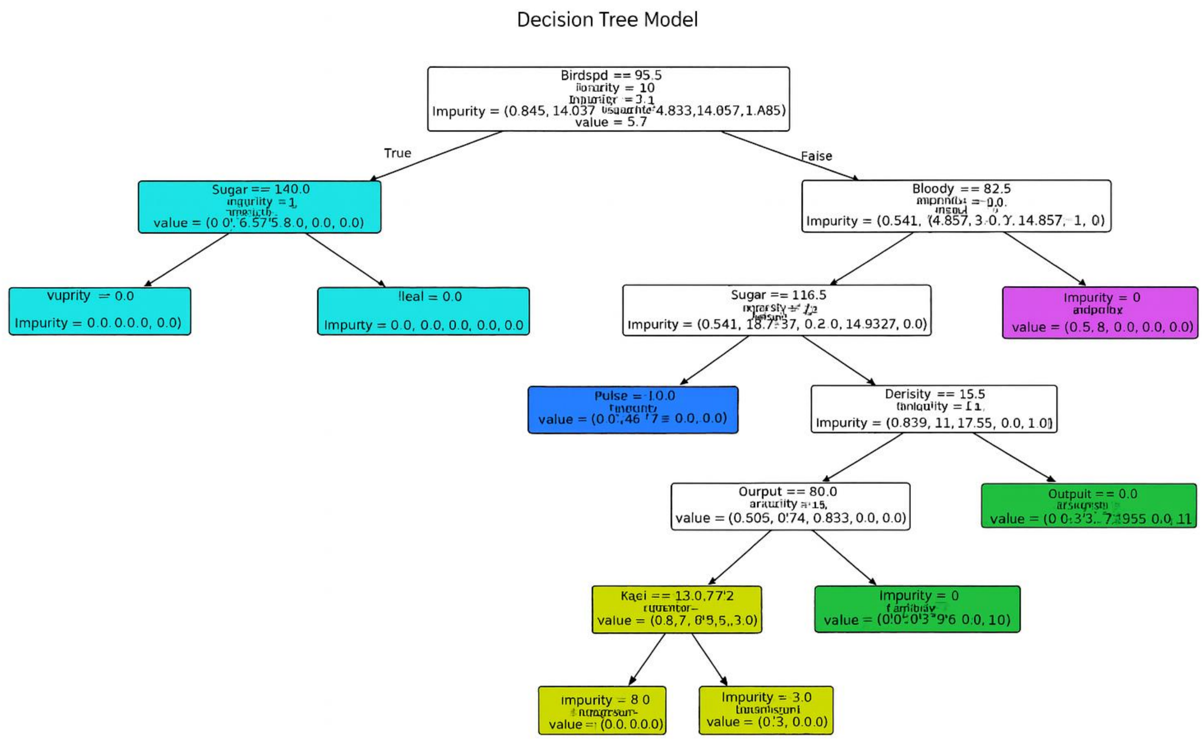

Figure 8. Decision Tree

This product integrates several vital sensors connected to an ESP32 microcontroller to monitor real-time heart health conditions. Before collecting sensor data, blood pressure is first checked using a digital *sphygmomanometer* as initial data. Data collection from the sensors takes approximately 3–5 minutes to ensure more accurate measurement results. The MAX30102 sensor measures heart rate (*BPM*) and blood oxygen levels (*SpO₂*) by placing it on the left index finger. The GY-906 (MLX90614) sensor is used to monitor body temperature by measuring the left index finger simultaneously. Meanwhile, the AD8232 sensor records simple heart wave signals, with electrodes attached directly to the body according to the recommended placement points.

Figure 9. Product Results

Data from each sensor is sent to the ESP32 to be processed and analyzed using the Decision Tree algorithm, which then generates risk classifications based on measured vital parameters. The results of this analysis are displayed directly in the Blynk application and automatically saved to a Google Spreadsheet as a digital archive. The Blynk display on *Ardenia Heartcare* is designed using several structured widgets to facilitate the monitoring of patient health data. It includes a Label widget to display the application title and connection status (*Online/Offline*), with icons for settings, notifications, and returning to the main page.

Figure 10. Blynk Display

Below that is an Input Text widget for entering the patient's name, a Selector or Dropdown widget for selecting gender, and a Stepper widget for setting the patient's age. The following section uses a Value Display widget to display the sensor reading parameters, including BPM, body temperature, SpO_2, systolic and diastolic blood pressure, and ECG values. A SuperChart widget displays a graph of the changing values to visualize the ECG signal in real-time. In addition, there is a special Label widget to display the risk classification results generated by the algorithm. At the bottom are two Button widgets: the *Calculate* button to process the data and the *SAVE* button to save the results to Google Spreadsheet.

4. CONCLUSION

Based on the research and testing results, this study successfully implemented a decision tree algorithm based on vital parameters such as BPM, SpO2, blood pressure, systolic and diastolic, and body temperature. The model achieved an accuracy of 85%, with an error rate of 15%. With limited data, 104 training and 27 test data can produce a learning curve of 0.87. From the research data, it can be concluded that the majority of targets with the risk class (SpO2 below normal) were 74 people, while the risk class (abnormal middle-aged women) had the smallest number, namely three people. Feature importance analysis shows that the SpO_2 and body temperature parameters play the most dominant role in model decision-making. The learning curve results show that the training accuracy value is stable in the range of 0.90–0.95, which indicates that the model can learn the training data patterns well, but the difference with the validation accuracy indicates potential overfitting. Based on the ROC Curve, most classes have an AUC value close to 0.5, equivalent to a random guess. The highest AUC value was obtained by the Risk (Abnormal adult male) class at 0.481, while the Risk (Abnormal body temperature) class had the lowest value at 0.250. This shows that the model's ability to distinguish between classes still needs to be improved, especially for minority classes.

ACKNOWLEDGMENTS

The author would like to thank all those who have provided support during this research process, especially the supervisors, fellow researchers, and institutions that have provided facilities and resources. This research would not have been completed successfully without their assistance and cooperation.

REFERENCES

Ayu Syahfitri. 2025. "Internet of Things (IoT), Sejarah, Teknologi, Dan Penerapannya." Uranus : Jurnal Ilmiah Teknik Elektro, Sains Dan Informatika 3(1):113–20. doi:10.61132/uranus.v3i1.667.

Dataset, Heart Disease, and Rough Neural Network. 2024. "Modeling Heart Disease Classification Using Rough Neural Network : A Data- Driven Approach To the Cleveland Heart." 7(2):135–45.

Hidayat, Rahmat, Yandiko Saputra Sy, Teguh Sujana, Mirdatul Husnah, Haris Tri Saputra, and Finanta Okmayura. 2024. "Implementasi Machine Learning Untuk Prediksi Penyakit Jantung Menggunakan Algoritma Support Vector Machine." BIOS : Jurnal Teknologi Informasi Dan Rekayasa Komputer 5(2):161–68. doi:10.37148/bios.v5i2.152.

Karamina, Raden Roro Dalilati Nabilah, and Ilham Ilham. 2025. "Smart Healthcare Dengan IoT: Implementasi Dan Tantangan Pada Sistem Kesehatan Berkelanjutan." Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI) 8(1):226–32. doi:10.32672/jnkti.v8i1.8413.

Mensah, George A., Valentin Fuster, and Christopher J. L. Murray. 2024. "Europe PMC Funders Group Global Burden of Cardiovascular Diseases and Risks , 1990-2022 Cardiovascular Disease in All Regions." 82(25):2350–2473. doi:10.1016/j.jacc.2023.11.007.Global.

Mikrokontroler Esp 32 Sebagai Alat Monitoring Pintu Berbasis Web. 2022. "Muhammad Nizam, Haris Yuana, Zunita Wulansari." Jurnal Mahasiswa Teknik Informatika) 6(2):767–72.

Nasrullah, Asmaul Husnah. 2021. "Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris." Jurnal Ilmiah Ilmu Komputer 7(2):45–51. doi:10.35329/jiik.v7i2.203.

Rashad, Ibnu, R. Rizal Isnanto, and Catur Edi Widodo. 2023. "Klasifikasi Penyakit Jantung Menggunakan Algoritma Analisis Diskriminan Linier." J. Sistem Info. Bisnis 13(1):29–36. doi:10.21456/vol13iss1pp29-36.

Ryfai, Denis Ahmad, Nurul Hidayat, and Edy Santoso. 2022. "Klasifikasi Tingkat Resiko Serangan Penyakit Jantung Menggunakan Metode K-Nearest Neighbor." Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer 6(10):4701–7. https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11662.

Selayanti, Nabilah, Shafira Amanda Putri, Mirechelin Kristanaya, Melinda Putri Azzahra, Muhammad Ghinan Navsih, and Kartika Maulida Hindrayani. 2024. "Penerapan Machine Learning Algoritma Random Forest Untuk Prediksi Penyakit Jantung." Prosiding Seminar Nasional Sains Data 4(1):895–906. doi:10.33005/senada.v4i1.376.