https://doi.org/10.48047/AFJBS.6.14.2024.5026-5037

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Cardio Care: A Data-Driven Approach for Heart Disease Prediction

Gorla Hemanth Reddy[1], Ravilla Girish Chandra[2], Sadineni Ajay[3], Tallapureddy Sai Venkat Reddy[4], Burra Vijay Babu[5]

[1]*Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P., India. Hemanthreddygorla7@gmail.com*

[2]*Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P., India. girishchandra00003@gmail.com*

[3]*Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P., India. Sadineniajay03@gmail.com*

[4]*Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P., India. Saivenkatt20021@gmail.com*

[5]*Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P., India. vijay_gemini@kluniversity.in*

Volume 6, Issue 14, Aug 2024

Received: 15 June 2024

Accepted: 25 July 2024

Published: 15 Aug 2024

doi: 10.48047/AFJBS.6.14.2024.5026-5037

Abstract - This research papers gives an analysation over the features of the heart disease analysis module that deals with a wide range of complications. Heart can be called as the most crucial and useful organ in the body of any organism especially in the human body. It plays a very important and purpose full role. Diagnosis of heart is very important and regular check-ups regarding heart and heart related diseases can be helpful in living a healthy life. To tackle such heart complications and to help people be cautious regarding their heart conditions there exists an essential urge of heart disease prediction system. Data sciences provides the noticeable support in predicting any type of the event which demands a structural training from the naturally occurring events. Data sciences and data science can be considered as the most popular and infamous platform for the sake of disease analysis, while the convenience of utilising it along with the computational domination in the array of modules that grabs the attention millions of patients all over globe, which also demands the several technical concerns in different platforms. In this research paper we will be analysing the accuracy of data sciences in predicting the heart diseases.

Keywords: Heart Disease, Cardio Care, Disease Analyses.

1. Introduction

Imagine a future where your everyday mobile device serves not only as a digital companion but also as a vigilant guardian of your health. Empowered by the insights gleaned from our model, it could discreetly alert you to potential heart-related issues, guiding you towards internationally recognized preventive measures. This isn't merely a vision of the future; it's the tangible promise of our research. Our hybrid model, acting as the conductor of this symphony of algorithms, surpasses the performance of its individual components, boasting an impressive specificity of 82.11% and a sensitivity of 91.47%. These numbers aren't just cold statistics; they represent lives preserved and families spared from the anguish of loss. This research heralds a paradigm shift in our approach to combating heart disease. It's not solely about diagnosis; it's about prevention empowering individuals to take proactive control of their health before the storm clouds gather. We envision a world where smart devices embedded with these intelligent systems seamlessly integrate into our daily lives, silently standing guard over our hearts. This endeavour transcends mere academic pursuit; it shines as a beacon of hope in the battle against the silent killer. Heart disease, often lurking in the background like a distant murmur, poses a significant threat to humanity. It's not a single adversary but a complex array of conditions, ranging from blocked arteries to irregular heartbeats, collectively endangering the core of human existence. This silent epidemic, frequently synonymous with the broader term "cardiovascular disease," exacts a heavy toll, claiming numerous lives and leaving countless others debilitated, fracturing families and straining healthcare systems.

But what if we could change this grim narrative? What if we could expose the silent killer before it unleashes its full force? This is the ambitious goal of our research: to create a data-driven shield, a predictive model harnessing the power of advanced algorithms, capable of anticipating the onset of heart disease before it becomes a dire threat. Picture a world where your everyday smartphone isn't just a digital tool but also a vigilant guardian of your health. Fuelled by our hybrid model, blending Naïve Bayes, Artificial Neural Networks, and Support Vector Machines, it could quietly alert you to potential health risks, nudging you towards healthier choices. Beyond merely identifying heart issues, our model focuses on prevention, empowering individuals to manage their health proactively. We envision smart devices seamlessly woven into daily life, serving as discreet monitors of vital signs and encouraging healthier lifestyles. This isn't just about convenience; it represents a paradigm shift in how we approach heart disease, moving from reactive treatments to proactive care.

The usage of the disease analysis systems has data analysis increased in the recent years, so far data analysis is the most famous and the popular analysis platform and it provides the convenience of using the analysis systems along with the computational convenience to handle the wide array of modules. The heart disease analysis platform provides us with the easy structure and demands the mark up languages and the technologies like data science and data analysis to function as desired. The movement that is discreetly designed functions to deliver a solitary information for the sake of the interface of the patient and a service which on a whole completes the assigned contextual work. The usage of the smart devices for the health monitoring day by day has increased rapidly and most of the world population already possess these facilities. In this data analysis and data-oriented world and every citizen with heart complications needs to be updated with the all the emergency information. The usage of heart diseases analysis has elevated exponentially in the very recent years. Prediction of heart disease can be considered as one of the most significant functions in the health care and pharma industry. A tiny mistake or the human error can cause sever complication while dealing with heart and heart diseases. Because of inefficient treatment and lazy diagnosis are causing a rapid growth in the heart complications and treatment failures day by day. Unstable and fluctuated heart decease and the uncertainty regarding its severity makes the prediction of heart decease a complicated and challenging issue to look after for the cardiological department. There are several services that cardiological organisations provide like health and heart monitoring which stands out as the top more priority for the families all over the planet. The task of heart decease prediction system is complicated as it demands a huge number of experts and specialised and the calls were made without any sort of certainty. Heart is the most essential organ so the care for heart is always the top priority in the health industry. Comparative study is highly needed in the field of the heart disease prediction. Most of the heart patients are dying all over the world due to the late recognition of the diseases or due to the lack of accuracy of severity of the disease so there is a critical need of more efficient methods and algorithms for predicting the heart diseases

Data sciences can be considered as one of the most efficient and useful technology in present days for the sake of testing and training. ML is the branch of the AI which can be taken into the consideration as a broad field of learning machines will be emulating the abilities of man. Data sciences modules are trained in a manner where they can learn how to utilise and process the data, so the combination of both the artificial intelligence and data sciences is called as the machine intelligence. As we know data sciences modules learns from the naturally occurring phenomenon and natural parameters. Taking this feature f data sciences into the consideration, we have taken the biological parameters as the testing data such as the body fat, cholesterol, blood pressure, gender, age and so on. Based up on these biological parameters a comparison was performed for the sake of accuracy of different data sciences algorithms. We have used four algorithms such as Knn, decision tree, SVM and linear regression. Based up on the calculations regarding the accuracy of these algorithms we have concluded the best algorithm heart disease prediction.

2. Literature survey

Predicting the future can be a gamble, especially when navigating complex situations. Imagine trying to map a treacherous mountain path blindfolded. That's where Yin Kia Chiam's 2019 research comes in, [3]shining a light on the tools and techniques needed for accurate, challenging predictions. Their mission? To audit popular data readiness features and frameworks, like the trusty compass and map for our mountain journey. Chiam delved into a rich toolbox, exploring diverse features and testing them alongside seven powerful classification techniques. Think of these techniques as skilled guides – kNN, decision trees, and neural networks, to name a few – each with their own strengths and weaknesses. [11]By meticulously comparing their performance, Chiam aimed to identify the optimal combination for navigating even the most intricate prediction landscapes. But accurate predictions rely on more than just the right tools. Just like a bare map needs context to be useful, reliable data is the bedrock of any prediction. [7]Enter T. Vivekanandan and N. Ch. Sriman Narayana Iyengar's 2017 research, focusing on heart disease diagnosis, a complex terrain where every clue matters. Their target? The vast array of features, akin to the intricate pieces on a game board, each potentially holding a key to unlocking the puzzle of heart health.

Vivekanandan and Iyengar recognized the importance of choosing the right pieces – features with true practical relevance that could guide diagnosis like skilled strategists. They emphasized "functional coverage," ensuring the chosen features addressed data collection issues, akin to securing a complete and accurate map before embarking on the prediction journey. To find the most effective "winning move" in this diagnostic game, the researchers employed a modified differential evolution (DE) algorithm. Imagine this algorithm as a sophisticated advisor, carefully assessing each feature and analysing its contribution to accurate diagnosis. By sifting through ten different methods, they sought the champion approach, the one that could consistently identify the correct diagnosis, the ultimate victory in this critical game. [11]Both Chiam and Vivekanandan & Iyengar's research highlight the intricate dance between data, features, and algorithms in tackling challenging predictions. They remind us that the path to accurate insights isn't a solitary trek but a collaborative effort, requiring the right tools, relevant data, and the wisdom to choose the optimal strategies. With their pioneering work, they pave the way for more reliable predictions across diverse fields, helping us navigate the complexities of the future with greater confidence and clarity.

[21]Bowel diseases, a spectrum of inflammatory conditions plaguing the digestive tract, remain shrouded in complexity. Demystifying them requires navigating a labyrinth of information, a daunting task exacerbated by the sheer volume and contradictions often found in raw datasets. This is where the intricate dance of data preprocessing takes centre stage. Imagine sifting through a mountain of unlabelled data, searching for nuggets of truth amidst the clutter. Studies like that of [12]V. Krishnaiah et al. (2014) shine a light on this process. They propose meticulously isolating and identifying relevant information like monitoring methodologies, aiming to build accurate predictive models for bowel disease patients. This approach resonates particularly with data structures within healthcare settings, where accurate diagnosis translates to effective treatment regimens. But the challenge lies not just in quantity, but also in quality. Raw data can be a double-edged sword, its abundance masking inconsistencies and redundancies. [4]This is where Lamia AbedNoor Muhammed's (2012) research comes in. They emphasize the crucial role of data selection and exclusion in preprocessing. By carefully choosing and representing informative symptoms and attributes, they pave the way for building refined models with enhanced predictive power. However, data alone is not enough. To navigate the labyrinth, we need sophisticated tools. Studies like that of Yin Kia Chiam (2019) shed light on the power of dimension reduction techniques like Most Forward Part Search (PCA). These techniques act as data sculptors, carving away irrelevancies and tailoring the information to specific needs. The result? Precise predictions that emerge from a streamlined landscape of meaningful data.

[13]But accuracy is not the only battleground. Clinical evaluations, often weighed down by the sheer size and ambiguity of data, can be susceptible to inaccuracies and even unnecessary patient testing. Here, research by T. Vivekanandan and N. Ch. Sriman Narayana Iyengar (2017) offers a potential solution. Their investigation into grouping techniques like Perceptron Multilayer (MLP) and back-spread learning algorithms proposes the inclusion of data constraints to identify characteristics that minimize unnecessary tests. This patient-centric approach seeks to streamline diagnostic processes, not only improving accuracy but also sparing individuals from undue discomfort. As we delve deeper into the intricate world of bowel disease,

research increasingly points towards the power of organized neural structures and advanced methodologies. [18]These sophisticated tools hold the promise of designing patient discovery systems, predicting with unparalleled precision, and ultimately paving the way for a future where bowel disease diagnosis is swift, accurate, and minimally invasive. The journey towards unravelling the mysteries of bowel disease is far from over. Yet, the insights gleaned from these and countless other studies provide a glimmer of hope. By mastering the art of data preprocessing, embracing advanced techniques, and prioritizing patient well-being, we inch closer to a future where bowel diseases are no longer hidden in the shadows, but brought to light, understood, and effectively managed.

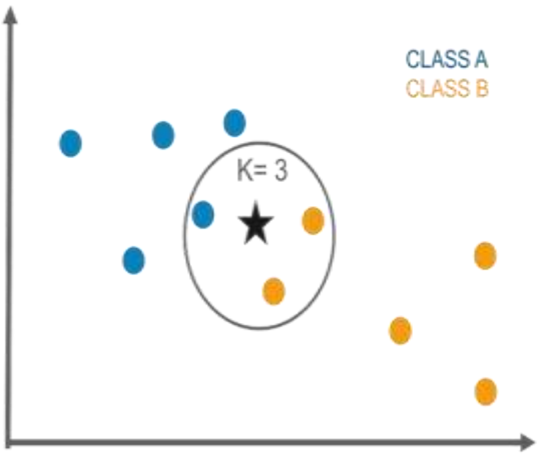
3. Methodology

This endeavour embarks on a data-driven journey, with Python as our trusted navigator and data science as our guiding compass. Our journey begins with a meticulously curated dataset sourced from the University of California, Irvine's online repository. Comprising 303 records and defined by 14 carefully selected attributes, this dataset forms the fertile ground for our investigation. These attributes, meticulously outlined in the accompanying table, serve as the breadcrumbs guiding our exploration into the heart of our research. The decision to employ Python as our platform was far from arbitrary. Its adaptability and widespread adoption within the data science community make it the ideal vehicle for traversing the intricate landscape of our dataset. Leveraging Python's robust toolkit, we seamlessly integrate data analysis into our project's framework, systematically uncovering patterns, correlations, and insights that steer us towards informed decisions.

Our dataset isn't just a reservoir of information; it serves as the driving force behind our research, propelling us to the forefront of data-driven exploration. Each record, each attribute, acts as a vital puzzle piece, meticulously assembled to reveal a clearer and deeper understanding of the subject matter at hand. With Python at the helm, guided by the principles of data science, and fuelled by our meticulously crafted dataset, our project is poised to chart new territories and offer invaluable insights. It's a testament to the potent synergy between code and information, driving us towards innovative discoveries and transformative insights.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

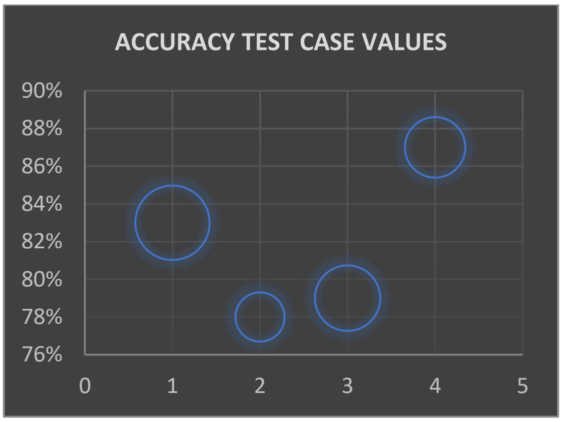
Imagine age as a winding road, some smooth stretches, others riddled with bumps. This, in essence, is what we're exploring: the landscape of age in relation to heart disease. To navigate this terrain, we've marked two paths: "target = 1" for those navigating the bumpy road of heart disease, "target = 0" for those cruising on the smooth one. This isn't just about simplification; it's about clarity. By splitting the data based on heart disease presence, we can see which age groups are more likely to encounter its bumps. We can pinpoint the steep hills, the stretches where the risk of taking the "target = 1" path significantly increases. This targeted approach lets us understand age's role in heart disease with laser focus, painting a detailed picture of vulnerable age groups and potential risk factors. The insights gleaned from this analysis won't be confined to statistics and graphs. They'll translate into action, informing preventive measures and healthcare strategies tailored to specific age groups. Knowing which stretches of the age road are riddled with heart disease bumps empowers us to build better roadmaps for a healthier future.



Imagine you're exploring a vast wilderness of data, seeking to uncover its secrets and navigate its twists and turns. Each data sciences algorithm is like a different guide, offering unique perspectives and approaches to help you on your journey. K-Nearest Neighbours (KNN) is your trusty scout, relying on the company of nearby data points to guide its predictions. Like a compass pointing towards familiar terrain, KNN looks at the traits shared by the closest "k" neighbours of a new data point to predict its likely path. This method works wonders when spatial or contextual similarities lead the way, making it perfect for uncovering hidden patterns in data that unfolds like a map. Decision Trees (DT), on the other hand, weave a branching narrative through the data landscape. They construct a map, carefully splitting data based on key features until each path leads to a distinct outcome. Their strength lies in their transparency – every fork in the road, every decision rule, is laid bare. This makes DT ideal for understanding the "why" behind predictions, offering insights into the inner workings of the data. Then there's Logistic Regression, taking a probabilistic approach to prediction. Like an expert odds-maker, Logistic Regression delves into the dance between variables, calculating the likelihood of an event happening based on past observations. This method excels in binary scenarios, confidently predicting outcomes like "yes" or "no" with a numerical confidence score. It's akin to understanding the weather forecast – providing a quantitative glimpse into the future based on past whispers. Choosing the right guide depends on the terrain you're traversing. KNN shines when nearby neighbours hold predictive power, DT when transparency is paramount, and Logistic Regression when probabilities paint the picture. By mastering these diverse tools, data scientists become skilled cartographers, adept at navigating the intricate labyrinths of data and revealing the hidden paths to understanding and informed decision-making.

Gaussian naive Bayes, within the realm of data sciences and artificial intelligence, represents a subset of elementary probabilistic classifiers. These classifiers hinge on the application of the Bayes hypothesis, incorporating a "naive" assumption of independence among attributes. This assumption simplifies the modeling process, making it computationally efficient and well-suited for various classification tasks. Gaussian naive Bayes leverages strong independence assumptions to compute probabilities, making it a pragmatic choice for scenarios where the naive assumption aligns with the underlying data distribution.

Support Vector Machines (SVM) stand out as one of the most prevalent artificial intelligence algorithms. Renowned for their efficiency, SVMs are an optimal choice for classification tasks, especially when slight modifications are needed. SVMs excel in creating an effective decision boundary in high-dimensional spaces, making them versatile tools in diverse applications. Random Forests emerge as a comprehensive learning technique in data sciences, employed for tasks such as classification and regression. During the training phase, numerous decision trees are generated, and the class with the highest frequency (for classification) or the average prediction (for regression) across these individual trees is determined as the final outcome. This ensemble approach enhances predictive accuracy and mitigates overfitting, demonstrating the versatility and effectiveness of random forests in handling complex datasets. Gaussian naive Bayes, Support Vector Machines, and Random Forests contribute to the diverse landscape of data sciences algorithms. Each algorithm addresses specific aspects of classification and prediction, with Gaussian naive Bayes relying on independence assumptions, Support Vector Machines excelling in efficiency with slight modifications, and Random Forests utilizing ensemble learning for enhanced accuracy. The selection among these algorithms depends on the nature of the data and the objectives of the analysis, reflecting the adaptability and applicability of data sciences methodologies in diverse scenarios



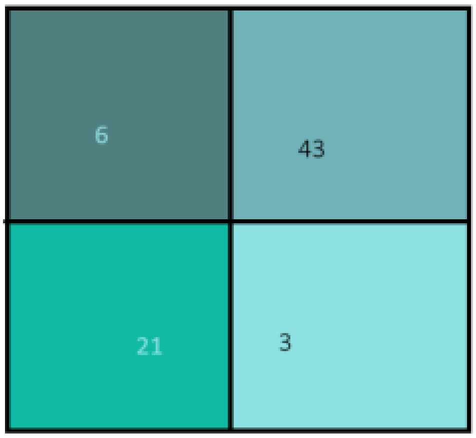
In the above graph shows us the test cases vs possible accuracy outcomes

4. Results

The confusion matrix serves as a visual aid, showcasing both the correct and incorrect predictions made by a classifier. It provides a comprehensive view of the classifier's performance by tallying the number of True Positives (TP) and True Negatives (TN), which represent accurately classified entries. To gain a deeper understanding, it's beneficial to examine the confusion matrices for each classifier individually. These matrices offer valuable insights into how well each classifier performs, highlighting instances of true positives, true negatives, false positives, and false negatives. Analysing these elements allows for a thorough evaluation of the classifier's ability to correctly identify positive and negative instances, providing a nuanced assessment of its predictive prowess. By collectively analysing the confusion matrices, we can fine-tune and enhance classifier performance, leading to improved accuracy and reliability in predictions within the specific context under consideration. This process is essential for refining the classifiers and ensuring their effectiveness in real-world applications.

ALGORITHM	ACCURACY
SVM	83%
Decision tree	78%
Logistic regression	79%
Knn	87%

The below confusion is the overall confusion matrix



References

[1] Verma A, Kaur I, Arora N. Comparative analysis of information extraction techniques for data mining. Indian Journal of Science and Technology. 2016 Mar; 9(11). DOI: 10.17485/ijst/2016/v9i11/80464.

[2] Chaurasia V. Early prediction of heart diseases using data mining techniques. Caribbean Journal of Science and Technology. 2013; 1:208–17.

[3] Alzahani SM, Althopity A, Alghamdi A, et al. An overview of data mining techniques applied for heart disease diagnosis and prediction. Lecture Notes on Information Theory. 2014; 2(4):310–5.

[4] Swathi P, Yogish HK, Sreeraj RS. Predictive data mining procedures for the prediction of coronary artery disease. International Journal of Emerging Technology and Advanced Enginereing. 2015, 5(2):339–42.

[5] Liu X, Lu R, Ma J, Chen L. Privacy-preserving patient-centric clinical decision support system on naïve bayesian classification. IEEE Journal of Biomedical and Health Informatics. 2016; 20(2):655–88.

[6] Patil RR. Heart disease prediction system using naïve bayes and jelinek-mercer smoothing. International Journal of Advanced Research in Computer Science and Communication Engineering. 2014; 3(5):6787–9.

[7] Pattekari SA, Parveen A. Prediction system for heart disease using naive bayes. International Journal of Advanced Computer and Mathematical Sciences. 2012; 3(3):290–4.

[8] Komal G, Vekariya V. Novel approach for heart disease prediction using decision tree algorithm. International Journal of Innovative Research in Computer and Communication Engineering. 2015; 3(11):11544–1.

[9] Shouman M, Turner T, Stocker R. Using decision tree for diagnosing heart disease patients. Proceedings of the 9th Australasian Data Mining Conference (AusDM'11); Ballarat, Australia. 2011. p. 23–30.

[10] Rani KU. Analysis of heart diseases dataset using neural network approach. IJDKP. 2011; 1(5):1–8.

[11] Majali J, Niranjan R, Phatak V, Tadakhe O. Data mining techniques for diagnosis and prognosis of cancer. International Journal of Advanced Research in Computer and Communication Engineering. 2015; 4(3):613–6.

[12] Iyer A, Jeyalatha S, Sumblay R. Diagnosis of diabetes using classification mining techniques. IJDKP. 2015; 5(1):1–14.

[13] V. Manikantan & S.Latha,"Predicting the Analysis of Heart Disease Symptoms Using Medicinal Data Mining Methods", International Journal on Advanced Computer Theory and Engineering, Volume-2, Issue-2, pp.5-10, 2013.

[14] Dr.A.V.Senthil Kumar, "Heart Disease Prediction Using Data Mining preprocessing and Hierarchical Clustering", International Journal of Advanced Trends in Computer Science and Engineering, Volume-4, No.6, pp.07-18, 2015.

[15] Uma.K, M.Hanumathappa, "Heart Disease Prediction Using Classification Techniques with Feature Selection Method", Adarsh Journal of Information Technology, Volume-5, Issue-2, pp.22-29, 2016

[16] Himanshu Sharma, M.A.Rizvi, "Prediction of Heart Disease using Data sciences Algorithms:A Survey",International Journal on Recent and Innovation Trends in Computing and Communication,Volume5,Issue-8,pp.99-104, 2017.

[17] S.Suguna, Sakthi Sakunthala.N ,S.Sanjana, S.S.Sanjhana, "A Survey on Prediction of Heart Disease using Big data Algorithms", International Journal of Advanced Research in Computer Engineering & Technology,Volume-6,Issue-3,pp.371-378,2017.

[18] A. L. Bui, T. B. Horwich, and G. C. Fonarow, "Epidemiology and risk profile of heart failure," Nature Reviews Cardiology, vol. 8, no. 1, pp. 30–41, 2011.

[19] J.Mourão Miranda, A.L.W. Bokde, C. Born, H. Hampel, and M. Stetter, "Classifying brain states and determining the discriminating activation patterns: support vector machine on functionalMRIdata,"NeuroImage,vol.28,no.4,pp.980–995, 2005.

[20] S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, "Innovative artificial neural networks-based decision support system for heart diseases diagnosis,"Journal of Intelligent Learning Systems and Applications, vol. 5, no. 3, pp. 176–183, 2013.

[21] Q. K. Al-Shayea, "Artificial neural networks in medical diagnosis," International Journal of Computer Science Issues, vol. 8, no. 2, pp. 150– 154, 2011.

[22] K. Vanisree and J. Singaraju, "Decision support system for congenital heart disease diagnosis based on signs and symptoms using neural networks," International Journal of Computer Applications, vol. 19, no. 6, pp. 6–12, 2011.

[23] Al Mamoon I, Sani AS, Islam AM, Yee OC, Kobayashi F, Komaki S (2013) A proposal of body implementable early heart attack detection system, 1-4.

[24] Patterson K (2016) Matthias Nahrendorf. Circ Res 119: 790-793. [13] Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive data mining for medical diagnosis: An overview of heart disease prediction. International Journal of Computer Applications, 17(8), 43-48.

[25] Masethe, H. D., & Masethe, M. A. (2014, October). Prediction of heart disease using classification algorithms. In Proceedings of the world congress on engineering and computer science (Vol. 2, pp. 22-24).

[26] A. Methaila, P. Kansal, H. Arya, and P. Kumar, "Early heart disease prediction using data mining techniques," in Proceedings of Computer Science & Information Technology (CCSIT-2014), vol. 24, pp. 53–59, Sydney, NSW, Australia, 2014.