ILKOM Jurnal Ilmiah Vol. 16, No. 1, April 2024, pp.100-107
Accredited 2nd by RISTEKBRIN No. 200/M/KPT/2020; E-ISSN 2548-7779 | P-ISSN 2087-1716

Research Article Open Access (CC–BY-SA)

Classification of Correlation Patterns Based on Electrocardiogram Data of Heart Defects Using the Pearson Correlation Coefficient Method

Sumiati [a,1,*]; Donny Fernando [a,2]; Hamonangan Iman Hasoloan [a,3]; Marlia Purnamasari [a,4]
[a] *Universitas Serang Raya, Jl. Raya Cilegon Drangong Serang - Banten and 42162, Indonesia*
[1] *sumiatiunsera82@gmail.com;* [2] *mr.donny2008@gmail.com;* [3] *vekky_repi@civitas.unas.ac.id;* [4] *Second Author Email;* [5] *marlia.unsera@gmail.com*
** Corresponding author*

Article history: Received October 08, 2023; Revised December 11, 2023; Accepted March 22, 2024; Available online April 26, 2024

Abstract

This study was conducted to map the relationship between a symptom and the type of heart disease, based on the results of the electrocardiogram medical record data. The purpose of this study was to apply a symptom correlation pattern based on electrocardiogram data of heart abnormalities. Where the results of this study produce values that determine symptoms that have a very close relationship with the type of heart disorder, and make an analysis to diagnose normal and abnormal heart disorders using the Pearson Correlation Coefficient (PCC) approach. The results show that the relationship between symptoms has a very strong relationship. dominant with normal heart defects is the relationship between AV conduction duration and other symptoms because the relationship between AV conduction duration and other symptoms has a very strong average level of association. symptoms also have a strong average level of association, while the relationship between other symptoms appears to have a moderate relationship and does not even have any relationship with someone who is identified as having a heart abnormality diagnosis (abnormal) and normal heart.

Keywords: Electrocardiogram; Diagnosis; Heart Disease; Pearson Correlation Coefficient; Symptoms.

Introduction

The number one death every year is caused by heart disease [1], One way to detect heart abnormalities is with the help of an electrocardiogram (EKG). Heart defects are health disorders that severely affect the major blood vessels when they supply blood to the heart, where the Pearson Correlation Coefficient (PCC) produces values that determine symptoms that have a very close relationship. with a type of heart disorder.

Monitoring systems such as ECG for diagnosis and treatment are switching to wireless technology with the development of medical instrumentation technology that is used mobile and makes patients feel more comfortable [1]–[3]. One of the wireless-based devices in medical telemetry that might be developed is telecardiology which involves transmitting electrocardiogram (ECG) signals [4]–[6], in addition to using computational methods with a Machine Learning (PCC) approach to diagnose abnormalities such as predicting heart disease [7]–[10]. , Apart from that, researchers use the K-Means concept to diagnose heart abnormalities [11]–[13], classify heart disease using a naive Bayes approach [14]–[16], By utilizing smartphone technology to diagnose heart abnormalities, they are able to predict the risk of sudden heart attack . [17], ECG diagnosis uses the Pearson correlation coefficient approach [18], Several computational approaches are used to diagnose heart abnormalities using artificial neural network algorithms [19]–[22], Another approach that is widely used to diagnose heart abnormalities is the fuzzy logic approach [23]–[26].

This research uses the Pearson Correlation Coefficient (PCC) method to determine the relationship between two variables, where this correlation technique is related to interval, ordinal and nominal scale variables. Pearson correlation is one of the correlation tests used to determine the degree of closeness of the relationship between two variables that have an interval or ratio, have a normal distribution, and returns a coefficient value with a range between -1.0 and 1, where a positive value is 1, a value of -1. is a negative value, and a value of 0 is a value where there is no correlation. This research aims to apply symptom correlation patterns based on electrocardiogram data on heart

disorders, where the results of this research produce values that determine symptoms that have a very close relationship with heart type. The results of the study showed that the relationship between symptoms that had a very dominant relationship with normal heart abnormalities was the relationship between AV conduction duration and other symptoms because the relationship between AV conduction duration and other symptoms had a very strong average level of relationship, in addition to the relationship between right and left ventricular activation. (ventricular depolarization). with other symptoms. symptoms also had a strong average level of correlation, whereas other symptom associations appeared to have moderate or even no association with someone being identified as having a diagnosis of heart disease.

Pearson Correlation Coefficient (PCC)

$$T_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}((x_i - \bar{x})^2 . \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (1)$$

$(x_i - \bar{x}), \bar{x} = x$ bar is the average obtained from the variable x

$(y_i - \bar{y}), \bar{y} = y$ bar is the average obtained from the variable y

Electrocardiogram

Normal and Abnormal Heart Abnormalities

Normal heart and abnormal heart heart sounds can be checked through a stethoscope. Abnormal heart sounds can be caused by problems with the heart valves, infections in the heart, and heart rhythm disturbances. The types of sounds in the heart consist of friction rub such as paper friction can be caused by friction in the pericardial layer and are found in patients with pericarditis, murmurs that are similar to hissing sounds, and gallop sounds like the soles of a horse's hoof can be experienced by patients with heart failure.

The normal heart rate when not actively moving reaches 60-100 beats per minute. This condition is called a resting heart rate, while a normal heart rate when actively moving, the heart can beat faster to distribute more oxygen to the muscles of the body that are working. In addition, if the heart rate when running is much faster than the heart rate when walking, so that the heart rate during exercise can reach 160 beats per minute or more.

Method

Collect electrocardiogram medical record data and conduct interviews with heart specialists. The next process is to carry out the process of interpreting heart abnormalities (abnormal) and normal heart. The next step is to analyze the correlation pattern between heart disease (abnormal) and normal hearts. The step with the PCC approach is to rank the attributes of a defective (abnormal) heart, the results of ranking the attributes of a normal heart. The stages in the process of diagnosing heart disorders using the Pearson correlation approach consist of entering patient data, reading patient data, carrying out data analysis, carrying out the Pearson correlation test and that's it.

Data is a very important component in the success of research, it is related to the way data is collected, the data sources used. Based on the data source, the type of data used in this study is primary data. The data used in this study was taken from electrocardiogram medical records which consisted of HR value intervals, P-R values, QRS values, QT values, QTC values, AXIS values. , RV6 value, SV1 value and R+S value.

Results and Discussion

A. Abnormal Heart Disease Attribute Ranking Results

This research aimsThe aim of this research is to apply symptom correlation patterns based on electrocardiogram data on heart disorders using the Pearson correlation coefficient method. Where the results of this research produce values that determine symptoms that have a very close relationship with the type of heart abnormality, and create an analysis to diagnose normal and abnormal heart abnormalities using the Pearson correlation coefficient approach. One application of the Pearson correlation coefficient is to continuously increase the level of linear relationship between two variables. Where normal and abnormal heart defects have 9 attributes (symptoms), namely HR, P-R, QRS, QT, QTC, AXIS, RV6, SV1 and R+S, which later each attribute (symptom) is connected to other attributes (symptoms) so that can produce an r value which will later determine the relationship that most influences normal and abnormal heart defects.

This study produces symptoms that have a linear relationship to two continuous variables, where the electrocardiogram medical record data is 100 data and has nine attributes, namely Heart Rate, PR, activation of the right and left ventricles (ventricular depolarization)., duration of ventricular depolarization and repolarization, Corrected QT interval, Frontal plane and Horizontal Plane, RV6, SV1 and R+S, where each attribute (symptom) returns a value r, where the value of r determines the most influential relationship to heart defects.

Analysis of the pattern of heart abnormalities and normal heart symptoms was tested using the WEKA software, The test is carried out by looking for attribute rankings, so that it can be seen that the relationship that has the highest value compared to other relationships will then be grouped according to the attribute ranking of the correlation of symptoms to normal and abnormal heart disorders. After all attributes have their respective ranking order, then choose the attribute that has the highest ranking from other attributes, in this study the results of the attribute with the highest ranking value are the RV6 attribute against other attributes, to make it easier to see the ranking order of the RV6 attribute. **Figure 1** shows the ranking of the correlation of symptoms of heart abnormalities and normal heart.

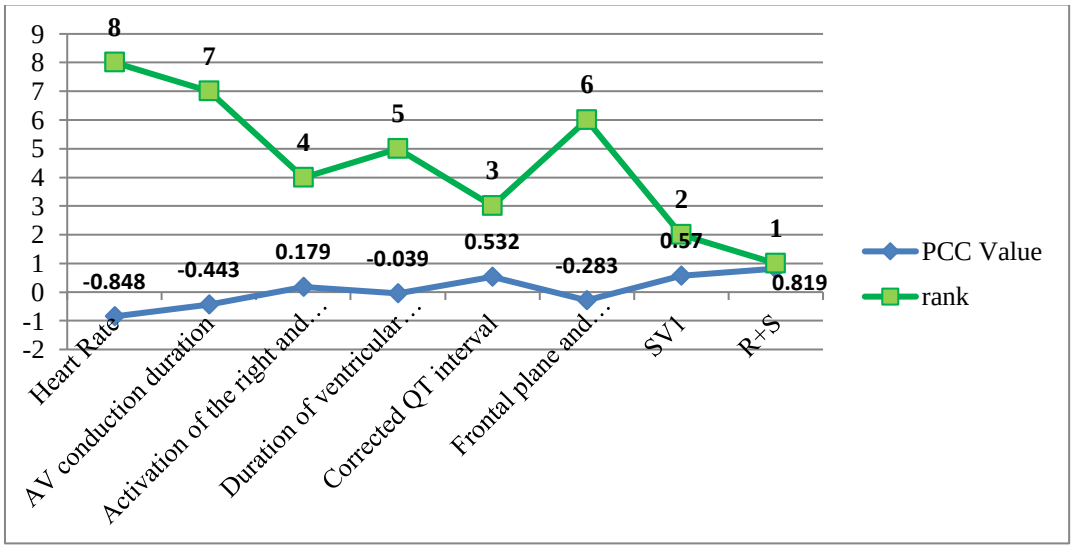


Figure 1. RV6 Rank Abnormal Heart Disease

B. Normal Heart Disease Attribute Ranking Results

After all the attributes have their respective ranking order, then choose the attribute that has the highest ranking of the other attributes, In this study, the results of the attribute with the highest ranking value are the AV conduction duration attributes against other attributes, to make it easier to see the ranking order of the AV conduction duration attributes.

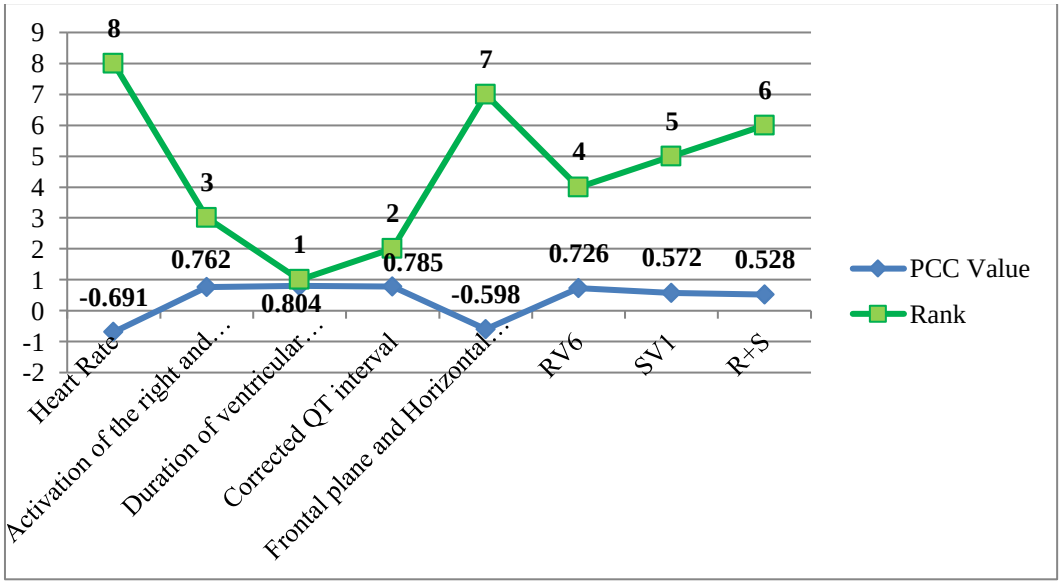


Figure 2. AV conduction duration Rank against Normal Heart

C. PCC Calculation Process for Medical Data of Heart Abnormal Patients (Abnormal)

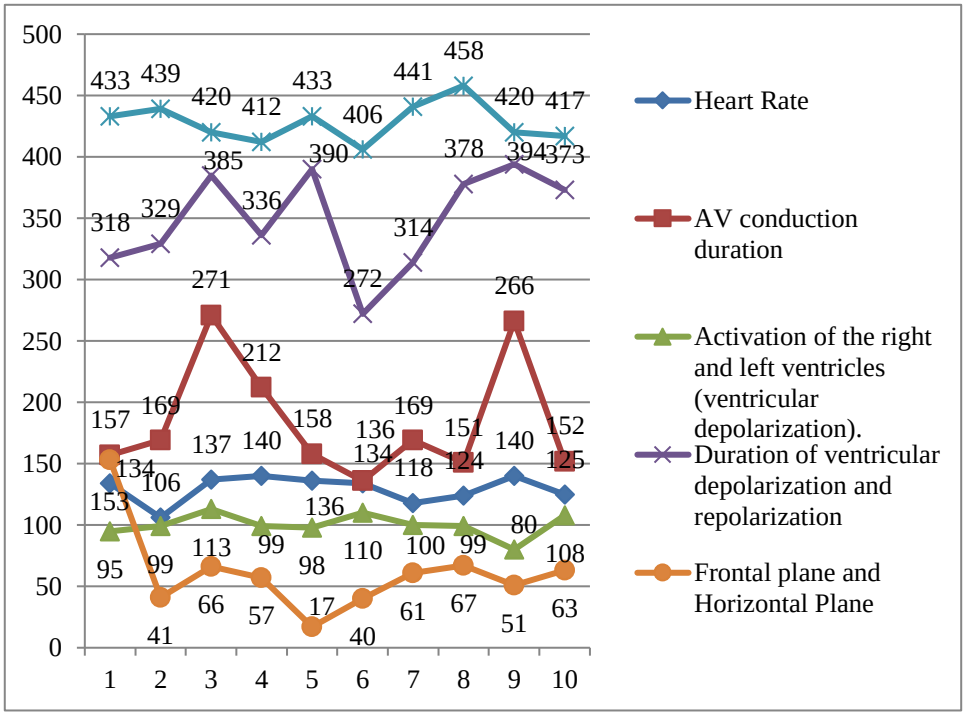


Figure 3. Medical Data of Abnormal Heart Abnormal Patients part 1

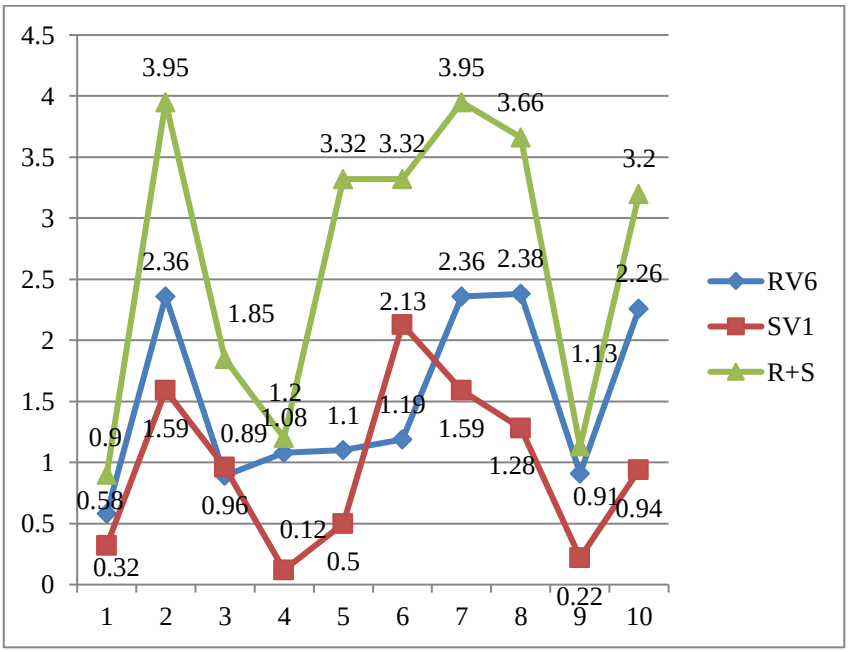


Figure 4. Medical Data of Abnormal Heart Abnormal Patients part 2

The calculation of the PCC value between Heart Rate and AV conduction duration can be shown as follows:

$$^{T}\text{Heart Rate } PR = \frac{\sum_{i=1}^{n} HR_i - \overline{\text{HR}})(PR_i - \overline{PR})}{\sqrt{\sum_{i=1}^{n} (HR_i - \overline{\text{HR}})^2 . \sum_{i=1}^{n} (PR_i - \overline{\text{PR}})^2}}$$

$$^{T}HR, PR = \frac{((HR_1-\overline{HR}).(PR_1-\overline{PR}))+((HR_2-\overline{HR}).(PR_2-\overline{PR}))+((HR_3-\overline{HR}).(PR_3-\overline{PR}))+((HR_4-\overline{HR}).(PR_4-\overline{PR}))+((HR_5-\overline{HR}).(PR_5-\overline{PR}))+((HR_6-\overline{HR}).(PR_6-\overline{PR}))+((HR_7-\overline{HR}).(PR_7-\overline{PR}))+((HR_8-\overline{HR}).(PR_8-\overline{PR}))+((HR_9-\overline{HR}).(PR_9-\overline{PR}))+((HR_{10}-\overline{HR}).(PR_{10}-\overline{PR}))}{\sqrt{\begin{matrix}(HR_1-\overline{HR})^2+(HR_2-\overline{HR})^2+(HR_3-\overline{HR})^2+(HR_4-\overline{HR})^2+(HR_5-\overline{HR})^2+(HR_6-\overline{HR})^2+(HR_7-\overline{HR})^2+\\(HR_8-\overline{HR})^2+(HR_9-\overline{HR})^2+(HR_{10}-\overline{HR})^2.(PR_1-\overline{PR})^2+(PR_2-\overline{PR})^2+(PR_3-\overline{PR})^2+(PR_4-\overline{PR})^2+(PR_5-\overline{PR})^2+(PR_6-\overline{PR})^2\\+(PR_7-\overline{PR})^2+(PR_8-\overline{PR})^2+(PR_9-\overline{PR})^2+(PR_{10}-\overline{PR})^2\end{matrix}}}$$

$$^{T}HR, PR = \frac{\begin{matrix}((134-129{,}40).(157-184{,}10))+((106-129{,}40).(169-184{,}10))+((137-129{,}40).(271-184{,}10))+((140-129{,}40).(212-184{,}10))\\+((136-129{,}40).(158-184{,}10))+((134-129{,}40).(136-184{,}10))+((118-129{,}40).(169-184{,}10))+\\((124-110{,}60).(151-184{,}10))+((140-129{,}40).(266-184{,}10))+((125-129{,}40).(152-184{,}10))\end{matrix}}{\sqrt{\begin{matrix}(134-129{,}40)^2+(106-129{,}40)^2+(137-129{,}40)^2+(140-129{,}40)^2+(136-129{,}40)^2+(134-129{,}40)^2+(118-129{,}40)^2+\\(124-110{,}60)^2+(140-129{,}40)^2+(125-129{,}40)^2.(157-184{,}10)^2+(169-184{,}10)^2+(271-184{,}1)^2+(212-184{,}10)^2+(158-184{,}10)^2\\+(136-184{,}10)^2+(169-184{,}10)^2+(151-184{,}10)^2\\+(266-184{,}10)^2+(152-184{,}10)^2\end{matrix}}}$$

$$HR, PR = \frac{2151.6}{\sqrt{(1094.4).(21348.9)}}$$

$$^{T}HR, PR = \frac{2151.6}{\sqrt{23364236.16}}$$

$$^{T}HR, PR = \frac{2151.6}{4833.656603}$$

$$^{T}HR, PR = 0.445$$

Table 1. Table of Symptoms of Normal Heart Abnormalities

Heart Rate	AV conduction duration	Activation of the right and left ventricles (ventricular depolarization).	Duration of ventricular depolarization and repolarization	Corrected QT interval	Frontal plane and Horizontal Plane	RV6	SV1	R+S
Heart Rate	1							
AV conduction duration	-0.69	1						
Activation of the right and left ventricles (ventricular depolarization).	-0.73	0.76	1					
Duration of ventricular depolarization and repolarization	-0.80	0.80	0.78	1				
Corrected QT interval	-0.59	0.78	0.59	0.59	1			
Frontal plane and Horizontal Plane	0.37	-0.60	-0.39	-0.52	-0.86	1		
RV6	-0.46	0.73	0.61	0.49	0.72	-0.66	1	
SV1	-0.33	0.57	0.62	0.41	0.33	-0.04	0.36	1
R+S	-0.60	0.53	0.65	0.49	0.61	-0.40	0.60	0.74

Based on the PCC value, it can be seen that the relationship between HR and almost all attributes (symptoms) does not have any relationship to a person's diagnosis when a normal heart defect is identified, however there is one relationship between HR and AXIS that is identified but has a moderate relationship.

The relationship between Heart Rate and almost all attributes (symptoms) does not have any relationship with a person's diagnosis when a heart abnormality is found, but there is a relationship between Heart Rate and Frontal plane and Horizontal Plane which is identified but has a moderate relationship. For the AV conduction duration R attribute (symptoms) it can be seen that the AV conduction duration to activation of the right and left ventricles (ventricular depolarization)., duration of ventricular depolarization and repolarization, Corrected QT interval, and RV6 has a very

strong relationship to the diagnosis of a person identified with normal heart abnormalities, where the AV conduction duration value to activation of the right and left ventricles (ventricular depolarization). is r 0.76 or 57.76%, AV conduction duration to duration of ventricular depolarization and repolarization is r 0.80 or 64%, AV conduction duration to Corrected QT interval is r 0.78 or 60.84% and AV conduction duration to RV6 is r 0.73 or 53.29%. Where two relationships of AV conduction duration symptoms to SV1 and R+S have a strong relationship to normal heart abnormalities and one relationship that does not have any relationship, namely AV conduction duration to Frontal plane and Horizontal Plane. The activation of the right and left ventricles (ventricular depolarization). attribute (symptom) has a strong relationship with the Corrected QT interval, RV6, SV1 and R+S attributes, where there is only one symptom relationship that appears to have a very strong relationship, namely the relationship between activation of the right and left ventricles (ventricular depolarization). and duration of ventricular depolarization and repolarization with an r value of 0.78 or 60.84% of normal heart defects and only has one relationship that does not have any relationship, namely activation of the right and left ventricles (ventricular depolarization). and Frontal plane and Horizontal Plane. In addition, the relationship between duration of ventricular depolarization and repolarization and Corrected QT interval symptoms only has a strong relationship with r 0.59 or 34.81% but has three moderate relationships, namely duration of ventricular depolarization and repolarization to RV6, SV1 and R+S and one relationship that does not have any relationship, namely duration of ventricular depolarization and repolarization to Frontal plane and Horizontal Plane. For the Corrected QT interval symptom relationship, there is only one very strong relationship, namely Corrected QT interval to RV6 with an r value of 0.72 or 51.84%, and only has one strong relationship, namely Corrected QT interval to R+S, and a moderate relationship, namely Corrected QT interval to SV1 and one relationship that does not have any relationship, namely Corrected QT interval against Frontal plane and Horizontal Plane. For Frontal plane and Horizontal Plane symptoms, it can be seen that all the relationships between A Frontal plane and Horizontal Plane XIS symptoms and other symptoms do not have any relationship to a person's diagnosis of normal heart abnormalities. For the symptoms shown by the RV6 and SV1 ranges, only the relationship between SV1 and R+S has a very strong relationship with an r-value of 0.74 or 54.76% of a person's diagnosis of normal heart defects. While the attributes (symptoms) of RV6 to SV1 only have a moderate relationship and the relationship of RV6 to R+S only appears to have a strong relationship.

The P-R relationship to other symptoms has a very strong average level of relationship, apart from that the QRS relationship to other symptoms also has a strong average level of relationship, while the relationship between other symptoms only appears to have a moderate relationship or even no relationship whatsoever to someone who is identified as having a diagnosis of a normal heart defect. It can be concluded that the relationship between symptoms that have a very dominant relationship with normal heart defects is the relationship between AV conduction duration and other symptoms because the relationship between AV conduction duration and other symptoms has a very strong average level of relationship. activation of the right and left ventricles (ventricular depolarization). associations with other symptoms also had a strong average association rate, whereas associations between other symptoms appeared to have a moderate or no relationship with someone identified as having the diagnosis.

Conclusion

Correlation pattern of symptoms based on electrocardiogram data on abnormal and normal heart abnormalities, starting with collecting data related to heart abnormalities and an EKG (Electrocardiogram) device, then managing the data through any factors related to abnormal and normal heart abnormalities and then determine correlation results. By using the Pearson correlation coefficient method, you can look for a level of correlation or relationship that is the strongest or has an influence on the identified symptoms of abnormal and normal heart abnormalities. By calculating the X and Y values and then looking for the average value of the X and Y values, you can find out the r value of a symptom's relationship to other symptoms. The Pearson correlation coefficient method can determine the correlation pattern of symptoms based on electrocardiogram data on abnormal and normal heart abnormalities. Because the Pearson correlation coefficient method is used to measure the strength and/or relationship of two variables.

This research was analyzed using WEKA, and only displays the ranking attribute for each symptom, so the researcher has to select again manually to find the highest ranking for each symptom attribute that has been analyzed using WEKA. In the WEKA application, patient data can only be processed using CSV or Arff format. So you have to create a CSV or Arff file first.

Acknowledgement

Thanks For LPPM Universitas Serang Raya.

References

[1] U. Varshney, Pervasive Healthcare Computing—EMR/EHR,Wireless and Health Monitoring.New York: Springer, 2009.

[2] S. Scalvini, G. Martinelli, D. Baratti, D. Domenighini, M. Benigno, L. Paletta, E. Zanelli, and A.Giordano, "Telecardiology: One-lead electrocardiogram monitoring and nurse triage in chronic heart failure," J. Telemedicine Telecare, vol. 11, Suppl. 1, pp. 18–20, Jul. 2005.

[3] A. Alesanco and J. Garcıa, "Clinical assessment of wireless ecg transmission in real-time cardiac telemonitoring," IEEE Trans. Inf. Technol. Biomed., vol. 14, no. 5, pp. 1144–1152, Sep. 2010. S. Sneha and U. Varshney, "A wireless ECG monitoring system for pervasive healthcare," Int. J. Electron. Healthcare, vol. 3, no. 1, pp. 32–50, 2007.

[4] A. Turnip, M. I. Rizgyawan, and D. E. Kusumandari, "Application of Support Vector MachineClassifier on Developed Wireless ECG System," International Journal of Pharma Medicine and Biological Sciences Vol. 5, No. 3, 2016.

[5] D. Cypher, N. Chevrollier, N. Montavont, and N. Golmie, "Prevailing over wires in healthcare environments: Benefits and challenges," IEEE Commun. Mag., vol. 44, no. 4, pp. 56–63, Apr. 2006.

[6] K. Kang, K.-J. Park, J.-J. Song, C.-H. Yoon, and L. Sha, "A medicalgrade wireless architecture for remote electrocardiography," IEEE Trans. Inf. Technol. Biomed., vol. 15, no. 2, pp. 260–267, Mar. 2011.

[7] R. Tao, S. Zhang, X. Huang et al., "Magnetocardiography-based ischemic heart disease detection and localization using machine learning methods," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 6, pp. 1658–1667, 2019.

[8] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, Article ID 81542, 2019.

[9] F. I. Alarsan and M. Younes, "Analysis and classification of heart diseases using heartbeat features and machine learning algorithms," *Journal of Big Data*, vol. 6, no. 1, p. 81, 2019.

[10] F. Z. Abdeldjouad, M. Brahami, and N. Matta, "A hybrid approach for heart disease diagnosis and prediction using machine learning techniques," in *The Impact of Digital Technologies on Public Health in Developed and Developing Countries*, M. Jmaiel, M. Mokhtari, B. Abdulrazak, H. Aloulou, and S. Kallel, Eds., Springer International Publishing, Manhattan, NY, USA, 2020.

[11] R. C. Ripan, I. H. Sarker, S. M. M. Hossain et al., "A data-driven heart disease prediction model through K-means clustering-based anomaly detection," *SN Computer Science*, vol. 2, no. 2, p. 112, 2021

[12] Sellappan Palaniappan, Rafiah Awang "Intelligent Heart Disease Prediction System Using Data Mining Techniques"Department of Information Technology Malaysia University of Science and Technology Block C, Kelana Square, Jalan SS7/26 Kelana Jaya, 47301 Petaling Jaya, Selangor, Malaysia

[13] Dr.Mohanraj, SubhaSuryaa, Sudha, Sarath Kumar, "Heart Disease Prediction using K Nearest Neighbour and K Means Clustering", International Journal of Advanced Engineering Research and Science (IJAERS) 2016.

[14] Shinde, R., Arjun, S., Patil, P., & Waghmare, J. (2015). An Intelligent Heart Disease Prediction System Using K-Means Clustering and Naïve Bayes Algorithm. IJCSIT) International Journal of Computer Science and Information Technologies, 6(1), 637-639.

[15] Mrs.G.Subbalakshmi, "Decision Support in Heart Disease Prediction System using Naive Bayes", Indian Journal of Computer Science and Engineering

[16] R. Bhuvaneswari and K. Kalaiselvi, Naive Bayesian Classification Approach in Healthcare Applications International Journal of Computer Science and Telecommunications, [Volume 3, Issue 1, January 2012].

[17] Raihan M, Mondal S, More A, Boni P K, and Sagor M O F 2017 Smartphone Based Heart Attack Risk Prediction System with Statistical Analysis and Data Mining Approaches Advances in Science, Technology and Engineering Systems Journal 2 No 3 pp 1815-1822

[18] Tb A Munandar, S Sumiati,and V Rosalina" Pattern of symptom correlation on type of heart disease using approach of pearson correlation coefficient" IOP Conf. Series: Materials Science and Engineering 830 (2020) 022086 IOP Publishing doi:10.1088/1757-899X/830/2/022086

[19] Sonali, B. M. and Priyanka, W., (2014), Research Paper on Basic of Artificial Neural Network, International Journal on Recent and Innovation Trends in Computing and Communication, 2(1),96-100

[20] Amato,F., López,A., María,E., Méndez,P., Vaňhara,P., Hampl,A., Havel,J.,(2013). Artificial neural networks in medical diagnosis, Journal of APPLIED BIOMEDICINE, 11,47-58

[21] Gharehchopogh,F.S., Khalifelu, Z.A., (2011)," Neural Network Application in Diagnosis of Patient: A Case Study", International Conference on Computer Networks and Information Technology , 245 – 249, http://dx.doi.org/10.1109/ICCNIT.2011.6020937 .

[22] Koushal, K., (2012). Artificial Neural Networks for Diagnosis of Kidney Stones Disease. I.J. Information Technology and Computer Science, 7(3), 20-25

[23] V. Krishnaiah, G. Narsimha, N.S. Chandra, Heart disease prediction system using data mining techniques and intelligent fuzzy approach: a review. Int. J. Comput. Appl. 136(2), 43–51 (2016)

[24] P.K. Anooj, Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules. J. King Saud Univ. Comput. Inf. Sci. 24(1), 27–40 (2012)

[25] D. Pal, K.M. Mandana, S. Pal, D. Sarkar, C. Chakraborty, Fuzzy expert system approach for coronary artery disease screening using clinical parameters. Knowl. Based Syst. 36, 162–174 (2012)

[26] V. Khatibi, G.A. Montazer, A fuzzy-evidential hybrid inference engine for coronary heart disease risk assessment. Expert Syst. Appl. 37(12), 8536–8542 (2010)