

Weighted Minkowski Similarity Method with CBR for Diagnosing Cardiovascular Disease

Edi Faizal¹

Department of Informatics Management
STMIK AKAKOM Yogyakarta
Yogyakarta, Indonesia

Hamdani Hamdani²

Department of Computer Science
Universitas Mulawarman
Samarinda, Indonesia

Abstract—This study implements Case-Based Reasoning (CBR) to make the early diagnosis of cardiovascular disease based on the calculation of the feature similarity of old cases. The features used to match old cases with new ones were age, gender, risk factors and symptoms. The diagnostic process was carried out by entering the case feature into the system, and then the system searched cases having similar features with the new case (retrieve). The level of similarity of each similar case was calculated using weighted Minkowski method. Cases with the highest level of similarity would be adopted as new case solutions. If the value of similarity was $<0,8$, the revision would be conducted by an expert. The tests result conducted by the expert showed that the system was able to perform the diagnosis correctly. The test results are performed on the sensitivity of 100% and specificity of 83,33%. Meanwhile, the accuracy of 95,83% and the error rate of 4,17% is so that this research is relevant enough to be implemented in the medical area.

Keywords—CBR; cardiovascular; similarity; weighted minkowski

I. INTRODUCTION

Weighted Minkowski similarity method Cardiovascular Disease (CVD) is the term for a series of heart and blood vessel disorders. World Health Organization data of 2012 shows that CVD is the number one cause of death in the world. In 2008 there were 17,3 million people died from CVD, these numbers represent 30% of the cause of death in the world. There were 7,3 million people died because of coronary heart disease and 6,2 million because of stroke [1].

On this issue, it is necessary for Diagnosing Cardiovascular Disease using Case-based Reasoning (CBR) approach with Weighted Minkowski Similarity. Many of the early systems attempted to apply pure Rule-Based Reasoning (RBR) as reasoning by logic in the expert system [2]. However, for broad and complex domains where knowledge cannot be represented by rules (i.e. IF-THEN), this pure rule-based system encounters several problems [3]. Due to the difficulty of the knowledge acquisition process, computer experts have tried to learn other problem solving methods known as CBR using lambda value analysis on the weighted Minkowski distance model [2].

The knowledge representation of CBR is a case base occurred previously. CBR uses a solution from an earlier case similar to the current case to solve the problem. The method that can be used to calculate similarity is weighted Minkowski [4]. If a new case has a resemblance to the old one, CBR will

reuse the old case solution as a recommendation for the new case solution. But if there is none match, CBR will do adaptation by retaining the new case into the case database, so the CBR knowledge will increase [2]. The more cases stored in the case base, the smarter the CBR system will be.

Based on the above facts, it is necessary to establish a system capable of diagnosing cardiovascular disease. The built system is an implementation of CBR in which the problems in new cases are solved by adapting solutions from old cases that have occurred and CBR is an important technique in artificial intelligence, which has been applied to various kinds of problems in a wide range of domains [3]. We use this weighted Minkowski similarity method because it is very good in our case for completing the diagnosis of CVD.

II. RELATED WORK

Several studies in the domain of cardiovascular have been conducted by [5] used a structured poly tree concept and directed acyclic graphical model (DAG) to predict all cases that could cause coronary heart disease. Tests showed that the applied concept was more accurate and efficient in predicting heart disease before the physical examination. In the same years [6] proposed a new algorithm to predict heart disease using CBR techniques. Meanwhile, most of the algorithms are based on Binary data only. The system was implemented using Java and had successfully predicted different levels of risk of heart attack effectively [6].

The application of CBR in the field of cardiovascular disease has been done by [7] by building case-based expert system prototype of heart disease diagnosis. While [8] used CBR in building a multimedia decision support system (MM-DSS) of heart disease diagnosis. [6] used 110 cases for 4 types of heart disease. The two retrieval methods were used namely induction and nearest-neighbour. It showed that the accuracy of using the nearest neighbour method is better than that of the induction method, i.e. 100% and 53%. Meanwhile, [8] medical multimedia based clinical decision support system for operational chronic lung diseases diagnosis and training with 97,36% Sensitivity, 97,77% Specificity, 96,85% positive predictive value (PPV) and 93,90% negative predictive value (NPV).

CBR for diagnosing heart failure in children done by [9]. The research conducted by [10] was for face recognition using 3 (three) different local features namely Manhattan distance, weighted angle distance and Minkowski distance. The results

showed that Minkowski distance provided better results in terms of time to recognize faces that were 0.46, 0.45 and 0.43 seconds for Minkowski, Weighted angle and Manhattan respectfully. [11] has developed a mobile cancer management system (MCSM) prototype to diagnose cancer patients. The system developed was a combination of CBR and CBIR with similarity measure using weighted Minkowski method. Based on 600 images of breast cancer radiology tested, it resulted in 90% accuracy.

Based on the explanation above, a great number of researches to diagnose cardiovascular disease have been conducted. In fact, the process of cardiovascular disease diagnosis needs to involve some risk factors, gender and age of patients to improve the accuracy of the diagnosis. Specifically, there has been no research conducted to diagnose the type of Acute Myocardial Infarction (I21) disease. Meanwhile, [12] conducted research on Individual risk prediction model for incident cardiovascular disease using the Bayesian approach.

Research in the application of CBR to make a diagnosis has also been conducted with various degrees of accuracy, while the application of Minkowski method has been performed for certain purposes with a fairly good level of accuracy. This research conducted in this paper applied CBR to diagnose type I21 cardiovascular disease. The diagnostic process involved symptoms, risk factors, age and gender of the patient. The calculation of similarity used weighted Minkowski method. The research was expected to generate a system capable of diagnosing cardiovascular disease, especially type I21 with a good level of accuracy.

Meanwhile, Minkowski central partition model by [13] for the pointer to a suitable distance exponent and consensus partitioning using developed clustering algorithm capable of computing feature weights. In [14], Minkowski metric for feature weighting and anomalous cluster is initializing in K-Means Clustering.

III. PROPOSED METHOD

The system input was in the form of risk factors and symptoms data of the patient's disease, and then the data were made into a case. There were two types of cases namely target case and source case. The source case was the case data entered into the system that served as knowledge for the system, while the target case was a new case data of which the solution to be sought.

The diagnostic process began with inserting patients' data, risk factors and symptoms experienced by the patient and then the similarities to the case stored were counted. Each feature had a certain value of weight obtained from the experts. The similarity between features was calculated using local similarity formula, and then calculated as a whole using global equality formula.

The calculation resulted in each case was sorted from the highest value to the lowest value. The highest value was the case most similar to the new case. The value of similarity ranged from 0 to 1 (in the percentage from 0% to 100%). If the value was smaller than the threshold value that was ≥ 0.8 , the solution of the case must be re-shared by the expert. The

system output was the name of the disease most similar to the new case.

A. Knowledge Acquisition

Case base would be formed from a collection of medical record data of cardiovascular disease inpatient of Dr. Sardjito public Hospital, Yogyakarta. The next stage was to make knowledge acquisition process to collect knowledge data from the knowledge source. The source of knowledge was obtained from an expert (cardiovascular disease specialist / SpJP). In addition to the expert, knowledge material was also derived from the literature related to the problem, such as books, journals, articles, etc.

B. Case Representation

The representation is intended to capture the essential properties problems and make that information accessible to the problem-solving procedure [3]. Case data obtained from medical records were stored in a case base. The collected cases were represented in the form of a frame. The frame contained the relation among the patient data, the illness, the risk factors and the symptoms of the case. Levels of confidence/trust were given on the relationships of these data so that the case for the CMB system could be made based on the representation in which problem space was the risk factor and the symptoms of the disease and solution space where the name of a disease.

Every risk factor and symptom has a weight that indicates the level of importance of the disease. The weight value ranges from 1 to 10 and the greater the value of the weight, the more important the risk factor or symptom determine the patient's disease. The level of confidence showing the sureness of the diagnosis of the expert is based on the risk factors and symptoms experienced by patients.

C. Indexing

The index on a record consists of two parts, search-key (value) and pointer. The search key is the value of a record while a pointer is index position of the search key. Case data searching in the retrieving process requires one or more search keys. In the development of CBR for Cardiovascular disease, two search keys have been developed namely risk code and symptom code.

D. Retrieval and Similarity

Retrieval is the core of the CBR – the process found in the case-base, the cases closest to the current case. The most commonly investigated retrieval techniques so far are the k-nearest neighbour, decision tree and its derivative. This technique uses a similarity metric to determine the size of similarity between cases [1]. In this study, the similarity method used referred to Equation (1) [10].

$$d(C_i, C_j) = \left(\frac{\sum_{k=1}^n w_k^r * |d(C_{ik}, C_{jk})|^r}{\sum_{k=1}^n w_k^r} \right)^{\frac{1}{r}} \quad (1)$$

With $d(C_i, C_j)$ is similarity value between case C_i (new case) and case C_j (old case), n is number of attributes in each case, k is individual attributes, ranging from 1 to n , w is weight given 1 to k attribute and r is Minkowski factor (positive integer).

The r value was the positive number ≥ 1 , (from 1 to infinite). The research presented in this paper used $r=3$. The previous research conducted by [11] showed that with the use of $r=3$ resulted in maximum accuracy.

The weighted of features in diagnosing cardiovascular disease was necessary because of the difference between particular features. Weight value was obtained from experts/cardiovascular disease specialists.

Due to the weight difference given to features for each case and the handling of new symptoms that may arise in the new case, the equation (1) introduced by [10] needs to be modified. Modifications to deal with similar problems have been carried out by [3] by adding the value of trust and handling of new symptoms as shown in equation (2).

$$T_k(T_i, S_i) = \text{Sim}(S, T) * T_k(S_i) * \frac{n(S_i, T_i)}{n(T_i)} \quad (2)$$

With $T_k(T_i, S_i)$ is similarity normalization with the level of trust, $T_k(S_i)$ is expert confidence level on a case in source case, $n(S_i, T_i)$ is number of symptoms of target case appear in source case and $n(T_i)$ is the number of symptoms in the target case [14].

The modification was made in equation (1) with reference to equation (2) so that the research conducted used equation (3).

$$E(C_i, C_j) = \left[\frac{\sum_{k=1}^n \omega_k^r * |d_k(C_{ik}, C_{jk})|^r}{\sum_{k=1}^n \omega_k^r} \right]^{1/r} * T(C_j) * \frac{n(C_i, C_j)}{n(C_i)} \quad (3)$$

Where,

- $E(C_i, C_j)$: The global similarity value between case of C_i (target case) and C_j (source case)
- ω_k : Weight value given to- k attribute
- $d_k(C_{ik}, C_{jk})$: Local similarity value between attribute C_i to- k and attribute C_j to- k
- r : Minkowski factor (positive integer)
- $T(C_j)$: Percentage of case trust level in case base
- $n(C_i, C_j)$: Number of new case attributes (C_i) and appear in old case (C_j)
- $n(C_i)$: Number of attributes in new case (C_i)

The similarity of each aspect in two cases is computed by a particular local similarity function. The local similarity values are aggregated by means of a sum of weighted aspects [15]. Local similarities are divided into two types namely symbolic and numeric. Features involved in symbolic is symptoms while in numeric are age, gender, smoking habit, body weight and so forth. The symbolic feature was calculated by using an equation (4) [12].

$$d_k(C_{ik}, C_{jk}) = \begin{cases} 1, & \text{if } C_{ik} = C_{jk} \\ 0, & \text{if } C_{ik} \neq C_{jk} \end{cases} \quad (4)$$

The numeric feature was calculated by using an equation (5) [13].

$$d_k(C_{ik}, C_{jk}) = \begin{cases} 1, & \text{if } C_{ik} = C_{jk} \\ 0, & \text{if } C_{ik} \vee C_{jk} = 0 \vee C_{ik} \vee C_{jk} = \perp \\ \frac{\min(C_{ik}, C_{jk})}{\max(C_{ik}, C_{jk})}, & \text{else} \end{cases} \quad (5)$$

If the similarity level is high, the case will be reused, in which the old case solution will be reused as a new case solution. If there is no case that meets the threshold value, the expert needs to give a conclusion to the new case.

E. Case Revision

The case revision is the part of system adaptation performed by an expert. The expert would revise the name of the disease and the level of confidence of the disease as the result of the diagnosis which has similarities lower than 0,8. After being revised, the case becomes a new case base.

F. System Implementation

The system is divided into 2 categories based on user types namely expert and paramedic. Each category of the user has access to a system with different facilities. Expert administration has access to add new users to the system, enter knowledge data, enter and revise cases as the result of a diagnosis and to diagnose them. Users with paramedic type have access to input patients' data, diagnose new cases and store new cases.

G. System Assessment

System assessment is carried out by performing diagnostic tests to measure the system's ability to detect disease or exclude a person without the disease. In [2] explained that sensitivity and specificity are used to determine the accuracy of diagnostic tests. Predictive values can be used to estimate disease probabilities, but positive predictive values and negative predictive values vary according to the prevalence of disease.

The analysis was conducted by using 4 parameters namely TP, FP, TN and FN and then they subsequently were used in calculating sensitivity, specificity, PPV and NPV. Calculation of the values using equations (6), (7), (8), (9) [16].

$$\text{sensitivity} = \left[\frac{TP}{TP+FN} \right] * 100\% \quad (6)$$

$$\text{specificity} = \left[\frac{TN}{TN + FP} \right] * 100\% \quad (7)$$

$$\text{PPV} = \left[\frac{TP}{TP + FP} \right] * 100\% \quad (8)$$

$$\text{NPV} = \left[\frac{TN}{TN + FN} \right] * 100\% \quad (9)$$

$$\text{Accuracy} = \text{sensitivity} \frac{P}{(P + N)} + \text{specificity} \frac{N}{(P + N)} \quad (10)$$

$$\text{Error Rate} = \frac{FP + FN}{(P + N)} \times 100\% \quad (11)$$

With TP (True Positive) is Positive diagnosis results for positive data samples, TN (True Negative) is Negative diagnosis Results for negative data samples, FP (False Positive) is Positive diagnosis results for negative data samples, FN (False Negative) is Negative diagnosis results for positive data samples, P is Total of positive diagnosis results and T is the total of negative diagnosis results. These values will appear in the Confusion matrix.

According to [16], the Confusion matrix is a useful way to analyze how well the system recognizes the tuples of different classes. TP and TN provide information when the system is correct, while FP and FN notify when the system is incorrect. Sensitivity and specificity can be used for the classification of accuracy. Sensitivity can be designated as true positives (recognition) rate (the proportion of correctly identified positive tuples). While specificity is true negatives rate (the proportion of the correctly identified negative tuples). The function of sensitivity and specificity can be used to show the accuracy level by equation (10) and the level of the system error rate can also be calculated by equation (11).

IV. RESULT AND DISCUSSION

A. Case Base Filling Process

The initial stage of the use of the CBR system was the preparation and filling of the base case. The case data inputted in the case base were medical records of inpatients obtained from the medical records installation of Dr. Sardjito Public Hospital, Yogyakarta. There were 126 cases with 74 symptoms and 9 types of risk factors that of class I21 disease (Acute Myocardial Infarction).

Symptoms and risk factors have a weight that indicates the level of importance of the symptoms or the risk factors. The weight of symptoms or risk factors were obtained based on expert data ranging from 1 to 10. Before filling the case base, users must first input patients' data, disease data, Symptom data and risk factor data into the system.

B. Diagnostic Process

Generally, the process of diagnosis of cardiovascular disease can be performed by doctors in several ways. The first way is to consider the risk factors and symptoms (felt by the patient). It is known as anamnesis. Another way is to perform laboratory tests to ensure the diagnosis. In the CBR system, the system performs the diagnostic process by anamnesis.

The diagnostic process began with selecting patient data to be diagnosed then entering symptoms data and risk factors felt by a patient into the system by utilizing the input facilities provided. Having all data entered, the system would perform retrieve process and calculate the similarity level between the new case and the case in the case base using weighted Minkowski.

Each case was calculated based on 4 components namely symptoms, risk factors, gender and age. Symptoms were assessed on the basis of the appearance of a symptom in a case. If a certain symptom appeared then it was valued 1 and 0 otherwise. Gender was categorized into male and female, while patient age was grouped in 6 age ranges. Several risk factors were assessed by their appearance in a case (such as symptom assessment), but there were two risk factors calculated based on a certain range namely disease history and smoking.

For example, there was a case in the case base as shown in Table 1. The user diagnosed a new patient with data entered into the system as shown in Figure 1. Based on the case example, the system performed the process of calculating both cases after the user had clicked the Diagnose Result button.

The process of calculating the local similarity in the case was divided into 4 (four) sections namely of age, gender, risk factors and symptoms. The calculation of age and risk factors used equation (5), while the proximity of sex and symptoms used equation (4), and the global similarities were calculated by using equation (3).

TABLE I. CASE BASE SAMPLE

New Case	Description	Value	Weight
Age	86	6	3
Sex	Female	0	2
Risk Fact:			
R009	Smoking < 10 stick per day	1	4
Symptoms:			
G004	Procedural cough	1	3
G014	White colored Sputum	1	2
G021	Cold sweat	1	5
G028	Nausea	1	2
G030	Vomit	1	4
G036	Left chest pain	1	7
G040	Chest pain at rest	1	6
G052	Heartburn	1	7
G062	Chest pain penetrating into the back	1	8
G067	Asphyxia	1	6
Disease I21.1, with 100% confidence level			

a) Local similarities

Age proximity: $(\text{Min}(5,6))/(\text{Max}(5,6)) = 5/6 = 0,83$.

Risk Factor proximity: $R004 = (\text{Min}(0,1))/(\text{Max}(0,1)) = 0/1 = 0$ and $R009 = (\text{Min}(1,2))/(\text{Max}(1,2)) = 1/2 = 0,5$.

Gender proximity = 0, due to the difference of old case and a new case. Symptoms proximity: Symptoms G004, G014, G021, G028, G030, G036, G040, G052, G062 are valued 1 because both cases have the symptoms. Symptom G062 is valued 0 because only old case (case base) has the symptom.

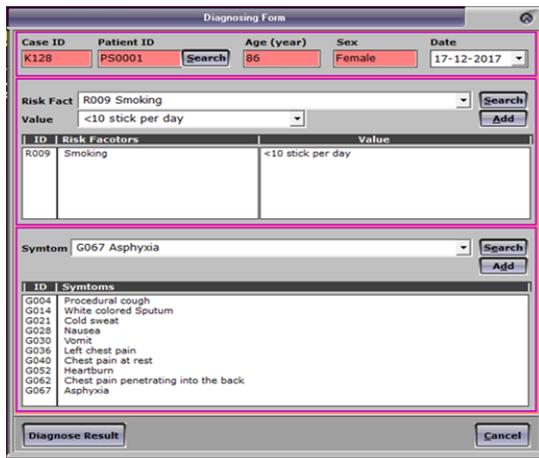


Fig 1. Diagnosis Process into the System.

1) *Global Similarities*: Based on the calculation of local similarities explained above, global similarities were calculated by using equation (3).

$$E(C_i, C_j) = \frac{\left[\begin{array}{l} (3 * 0,83)^3 + 0 + (4 * 0,5)^3 + \\ (3 * 1)^3 + \\ (2 * 1)^3 + (5 * 1)^3 + \\ (2 * 1)^3 + (4 * 1)^3 + \\ (7 * 1)^3 + (6 * 1)^3 + (7 * 1)^3 + \\ (8 * 1)^3 + 0 \end{array} \right]^{1/3}}{\left[\begin{array}{l} (3)^3 + (2)^3 + (4)^3 + (3)^3 + \\ (2)^3 + (5)^3 + (2)^3 + (4)^3 + (7)^3 + \\ (6)^3 + (7)^3 + (8)^3 + (6)^3 \end{array} \right]^{1/3}} * 100 * \frac{11}{13}$$

The similarity value of $E(C_i, C_j)$ was 0,80 obtained as the results of the calculation of the proximity of both cases described above. The system would display the results in the form Diagnosis Results. The output of the system was the name of the disease that the patient suffers from the highest level of similarity. The similarity value was between 0 and 1 and in a percentage form. The diagnose results would be retained into the system.

C. Case Revision Process

The process of case revision needs to be carried out if the system is not able to diagnose the disease correctly. The system is considered unsuccessful to diagnose the disease if the value of similarity is less than 0,8 (80%). Revision process can only be conducted by the user with access right as an expert.

D. System Assessment Process

The system assessment process was conducted in two ways namely the test conducted by experts and the test using a random data sample that was 30% of 126 cases or as many as 38 data. For test purposes, 10 cases were added as test data which was not a case of disease I21. The additional data were cases of heart failure (I50). The result of the assessment conducted by experts by using data samples based on their

knowledge, and then matched the results of the system with expert conclusions on the data tested.

System assessment was also performed by using medical record data. The value of experts' trust toward a case in the medical record data was 100% because it had been through a thorough observation and assessment. The assessment was conducted by using the data of case I21 and I50. The result of the system assessment using the data of case I21 showed that the highest similarity value was 0,99 and the lowest was 0,68. The average test result of the data of case I21 showed that the value was above the threshold of 95%. The test results using the data of case I50 showed that the highest similarity value was 0,77 and the lowest was 0,40 with the average value above the threshold of 100%.

The evaluation of the results of system assessment in diagnosing cardiovascular disease was conducted by calculating sensitivity, specificity, PPV, NPV, accuracy and error rate. Evaluation is important to determine whether the system built is feasible to be applied in diagnosing cardiovascular disease, especially for the type I21. The first stage to be done was to create a confusion matrix based on the value of similarity as the results of the system assessment, as shown in Table 2.

TABLE II. CONFUSION MATRIX OF SYSTEM TEST

		System Test		
		Case I.21	Case I.50	All Case
System test result	Similarity $\geq 0.8 (+)$	36 (TP)	0 (FN)	36 (P)
	Similarity $< 0.8 (-)$	2 (FP)	10 (TN)	12 (N)
	All Case	38	10	48 (P+N)

Confusion matrix shows 36 cases of disease I21 with similarity values $\geq 0,8$ and 2 cases with similarity values $<0,8$ that is on test 21 and 34 with similarity value of 0,79 and 0,68 respectively. The result obtained from the test using the case I50 is entirely $<0,8$. So that the level of a sensitivity, specificity, PPV, NPV, accuracy and error can be calculated by using equations (6), (7), (8), (9), (10) and equation (11).

$$\text{Sensitivity} = \frac{36}{(36 + 0)} * 100\% = 100\%$$

$$\text{Specificity} = \frac{10}{(10 + 2)} * 100\% = 83,33\%$$

$$\text{PPV} = \frac{36}{(36 + 2)} * 100\% = 94,74\%$$

$$\text{NPV} = \frac{10}{(10 + 0)} * 100\% = 100\%$$

$$\text{Accuracy} = 100\% * \left(\frac{36}{(36 + 12)} \right) + 83,33\% * \left(\frac{12}{(36 + 12)} \right) = 95,83\%$$

$$\text{Error rate} = \frac{(2 + 0)}{(36 + 12)} * 100\% = 4,17\%$$

The above calculation shows the percentage of the system's ability to recognize disease I21 correctly is 100% (sensitivity), the percentage of system ability to recognize disease which is not I21 correctly was 83,33% (specificity), positive predictive value was 94,74% (PPV), negative predictive value was 100% (NPV), And the accuracy was 95,83% with an error rate of 4,17%.

V. CONCLUSION

Based on the research and the results of the tests that have been conducted, it can be concluded that this study resulted in a case-based reasoning system with weighted Minkowski similarity calculation method used to perform early diagnosis of cardiovascular disease. This system performed a diagnostic process by taking into account the proximity between the case base and the target case based on the patient's condition (symptoms and risk factors), sex and age. The test results of the system for early diagnosis of cardiovascular disease using medical records of patients with disease I21 (based on case basis) and medical records of patients with I50 disease (not in accordance with the case basis), indicated that the system was able to recognize the disease I21 correctly (Sensitivity) of 100%, recognize non-I21 disease (specificity) of 83,33% with accuracy of 95,83% and error rate of 4,17%.

ACKNOWLEDGMENT

The research presented in this paper for conducted on the participation of various parties. Thank you for Dr Sardjito Public Hospital in Yogyakarta and Dr Irsad Andi Arso, M.Sc., Sp.PD., Sp.JP (K), who has been willing to be the resource for this research. We also expressed his gratitude to those who work in medical records installation of Dr Sardjito Public Hospital in Yogyakarta for the provision of medical record data as supports for this research.

REFERENCES

- [1] S. Wang, M. Petzold, J. Cao, Y. Zhang, and W. Wang, "Direct Medical Costs of Hospitalizations for Cardiovascular Trends and Projections," *Medicine (Baltimore)*, vol. 94, no. 20, pp. 1–8, 2015.
- [2] A. Labellapansa, A. Yulianti, and A. C. Reasoning, "Lambda Value Analysis on Weighted Minkowski Distance Model in CBR of Schizophrenia Type Diagnosis," in 4th International Conference on Information and Communication Technologies, 2016, vol. 4, no. c, pp. 1–4.
- [3] S. H. El-sappagh and M. Elmogy, "Case Based Reasoning: Case Representation Methodologies," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 11, pp. 192–208, 2015.
- [4] R. Carbó-dorca, "Molecular quantum similarity measures in Minkowski metric vector semispaces," *J. Math. Chem.*, vol. 44, no. 3, pp. 628–636, 2008.
- [5] M. A. E. Abbas, "Anticipating of Cardiovascular Heart Diseases using Computer based Poly Trees Model," *Int. J. Comput. Appl.*, vol. 59, no. 13, pp. 19–27, 2012.
- [6] K. C. Shekar, M. Tech, D. Ph, and M. Tech, "Improved Algorithm for Prediction of Heart Disease Using Case based Reasoning Technique on Non-Binary Datasets," *Int. J. Res. Comput. Commun. Technol.*, vol. 1, no. 7, pp. 420–424, 2012.
- [7] A. M. Salem, M. Roushdy, and R. A. Hodhod, "A Case Based ExpertSystem for Supporting Diagnosis of Heart Diseases," *AIML J.*, vol. 5, no. 1, pp. 33–39, 2005.
- [8] P. Pal and S. Tomar, "A Medical Multimedia based Clinical Decision Support System for Operational Chronic Lung Diseases Diagnosis and Training," *Int. J. Comput. Appl.*, vol. 49, no. 8, pp. 1–12, 2012.
- [9] A. Cardiology, "Case-Based Reasoning for Diagnosis Heart Failure in," *Austin Cardiol*, vol. 1, no. 1, pp. 1–5, 2016.
- [10] M. K. Rao, K. V. Swamy, K. Anithasheela, and B. C. Mohan, "Face recognition using different local features with different distance techniques," *Int. J. Comput. Sci. Eng. Inf. Technol.*, vol. 2, no. 1, pp. 67–74, 2012.
- [11] B. M. Elbagoury, A. M. Salem, and M. Roushdy, "A Hybrid Case-Based and Content-Based Retrieval Engine for Mobile Cancer Management System - MCMS A Hybrid Case-Based and Content-Based Retrieval Engine for Mobile Cancer Management System - MCMS," *Int. J. Bio-Medical Informatics e-Health*, vol. 1, no. May 2014, pp. 15–19, 2013.
- [12] Y. Liu, S. L. Chen, A. M. Yen, and H. Chen, "Individual risk prediction model for incident cardiovascular disease: A Bayesian clinical reasoning approach," *Int. J. Cardiol.*, vol. 167, no. 5, pp. 2008–2012, 2013.
- [13] R. Cordeiro, D. Amorim, A. Shestakov, B. Mirkin, and V. Makarenkov, "The Minkowski central partition as a pointer to a suitable distance exponent and consensus partitioning," vol. 67, pp. 62–72, 2017.
- [14] R. Cordeiro, D. Amorim, and B. Mirkin, "Minkowski metric , feature weighting and anomalous cluster initializing in K-Means clustering," *Pattern Recognit.*, vol. 45, no. 3, pp. 1061–1075, 2012.
- [15] C. Yang, X. Yu, Y. Liu, Y. Nie, and Y. Wang, "Neurocomputing Collaborative filtering with weighted opinion aspects," *Neurocomputing*, vol. 210, pp. 185–196, 2016.
- [16] A. K. Akobeng, "Understanding diagnostic tests 1 : sensitivity, specificity and predictive values," *Acta Pædiatrica*, vol. 96, pp. 338–341, 2007.