International Journal of Research in Engineering, Science and Management
Volume 4, Issue 6, June 2021
https://www.ijresm.com | ISSN (Online): 2581-5792

Self-Diagnosis of Cancer Using Case Base Reasoning Algorithm

Nadir Kamal Salih*
Department of Computer and Electrical Engineering, University of Buraimi, AL Buraimi, Oman

***Abstract*: Autonomic computing systems are similar to those in the autonomic nervous system of the human body. Autonomic computing is a system that can manage itself. This research discussed a problem, which is diagnosis of one of the most common diseases today, which is cancer. People can diagnose this disease through doctors but they can also have made a self-diagnosis for this disease without needed any doctors when they are in their home or anywhere, so they can save their time and also reducing the cost. We used an algorithm to self-diagnosis of cancer, which is case base reasoning. Case base reasoning is an automated reasoning and decision-making process whereby we solved new problems through the experiences we had accumulated in solving previous ones. So from Previous results we can self-diagnosis of cancer.**

***Keywords*: autonomic computing, case base reasoning, cancer.**

1. Introduction

Autonomic computing is a system that can manage itself by self-configuration, self-healing, self-optimizing and self-protection. Paul Horn, senior vice president of research for IBM, coined the term autonomic computing in 2001[1], [2]. According to Horn, the industry's focus on creating smaller, less expensive, and more powerful systems was fueling. Autonomic computing systems are work like the human body's autonomic nervous system. An autonomic computing system would control the functioning of computer applications and systems without user, in the same way that the autonomic nervous input from the system regulates body systems without conscious input individual [3], [4]. Autonomic computing systems are exclusion of any human involvement. In fact, meaning the one of an administrator, from a rather demanding and time-consuming task of a computer stymieing operation and maintenance, it being able to function without having to be overseeing. Such a system is then expected to self-manage, thus providing the end users with uninterrupted peak performance [5]. In other words, the system should observe the internal and external conditions, as well as software and hardware issues, and lake actions to address them properly [6], [7]. This may include, for example, the process of obtaining software updates, installing them, reconfiguring if necessary, running tests, and, potentially, reverting the previous software version as it may turn out inevitable and necessary in the case of errors [8], [9]. Most precisely, such functionality may be achieving with the following four key characteristic components of the concept of self-management, i.e. self-configuration, self-optimization, self-healing, and self-protection as depicted in figure 1.

Fig. 1. Autonomic computing systems

The important of autonomic computing is to create computing systems capable of managing themselves largely than they do today. With the nature of autonomy, reactivity, sociality and pro-activity, software agents are promising to make autonomic computing system a reality. The inexperience staff need the guidance from the experience staff to improve their diagnostic handling skills. We have a lot of contributions that we got:

1. A survey was conducted on the trends and developments of recent CBR medical systems.
2. A case-based reasoning system is developed to demonstrate that diabetes mellitus can only be diagnosed manually beforehand.
3. How an algorithm that matches similarities improves system performance.
4. Reduce the time taken to reach a decision especially in an emergency case.

The rest of the paper is structured as follows: Section 2 discusses the related work. Section 3 presents technique for proposed solution. Section 4 represents the solution model and case study. Section 5 demonstrates the analysis of result gained from experiments and section 6 include the conclusion and future work.

2. Related Work

We have studied the diagnosis of engineering systems,

*Corresponding author: nadir@uob.edu.om

Table 1
Similarity functions

Function	Function formula
Hamming similarity	$sim_{ij=}\frac{matches_{k-1}^{m}(P_{ik},P_{jk})}{m}$
Manhattan distance	$d_{ij}=\sum_{k=1}^{m}W_k\lvert P_{ik}-P_{jk}\rvert$
Euclidean distance	$d_{ij}=\sqrt{\sum_{k=1}^{m}(W_k\,(P_{ik}-P_{jk}))^2}$
Sim 1	$sim_{ij}=1-\max(IP_{ij},OP_{ij}),\ IP_{ij}=\max(\min(P_{ik},P_{jk})),OP_{ij}=\min(\max(P_{ik},P_{jk}))$
Sim 2	$sim_{ij}=1-t_1*t_2,\ t_1=\min(IP_{ij},1-OP_{ij}),\ t_2=\frac{IP_{ij}+(1-OP_{ij})}{2}$
Canberra distance	$d_{ij}=\sum_{k=1}^{m}\frac{\lvert P_{ik}-P_{jk}\rvert}{P_{ik}+P_{jk}}$
Bray-curtis distance	$d_{ij}=\frac{\sum_{k=1}^{m}\lvert P_{ik}-P_{jk}\rvert}{\sum_{k=1}^{m}P_{ik}+P_{jk}}$
Squared chord distance	$d_{ij}=\sum_{k=1}^{m}(\sqrt{P_{ik}}-\sqrt{P_{jk}})^2$
Squared chi- Squared distance	$d_{ij}=\sum_{k=1}^{m}\frac{(\lvert P_{ik}-P_{jk}\rvert)^2}{P_{ik}+P_{jk}}$
Jaccard similarity coefficient	$sim_{ij}=\frac{c_i\cap c_j}{c_i\cup c_j}$

including the use of test equipment and the relevant health and safety considerations as in following: The researchers talk in [10], [11] about self-diagnosis and its prevalence at this time. Where people can check up for a specific disease while they are at their home using pharmacy tools purchased online. Also talked about the damages of self-diagnosis in some cases patients may use Incorrect test and misinterpretation of results. The authors in [12], [13] they described a simple self-diagnosis tool for depression and using this tool to estimate the prevalence of dermatitis in France. The ability of the questionnaire to distinguish between infected and uninfected individuals proved. Using this tool for self-diagnosis of depression, this study provides a previously unprecedented look at the high prevalence of skin disease in elderly individuals [14]. The researchers mentioned in [15], [16] a proposed system of artificial preventive health care methods by analyzing lifestyle. It is a system expert and smart diagnostic system for medical diagnosis that can give the same effect as the help of real experts with a health examination assistant and scientific and objective knowledge appropriate to the age and environment of change, through which the patient's health status is analyzed through information and he suggests appropriate care for him before and after treatment [17].

Case - based reasoning (CBR) was one of the emerging models for designing intelligent systems [18]. Recovery of similar cases was a basic step in Case - based reasoning, and the similarity measure played a very important role in case retrieval. Sometimes the Case - based reasoning systems were called similarity searching systems, the most important characteristic of which was the effectiveness of the similarity measure used to quantify the degree of resemblance between a pair of cases [19]. Similarity was the general notion in Case-Based Reasoning. Also, the Similarity was always considered between problems not solutions of cases. Similarity functions could be used to get a group of the nearest neighbor solutions for current case problem [20]. That can lead to catch accuracy by different functions in various using. Here in table1 bellow can see example of this function are Hamming distance, Manhattan distance, Euclidean distance, Sim1, Sim2, Canberra distance, Bray–Curtis distance, Squared Chord distance, Squared Chi-Squared distance and Jaccard similarity function Formula Hamming similarity.

3. Solution Model

When studying cancer diseases, it was found that breast cancer is the most prevalent in recent time. Breast cancer can be known as develops from breast tissue. Signs of breast cancer may include a lump in the breast, a change in breast form, skin dimpling, fluid from the nipple, a newly inverted nipple or a red or scaly skin patch. Bone pain, swollen lymph nodes, shortness of breath, or yellow skin may occur in those with distant disease spread. Due to the high incidence of breast cancer, applications or devices must be developed to diagnose this disease and this helps treat it early. We used CBR to diagnose breast cancer for this purpose and this software makes it easier for the patient to know whether or not he has breast cancer. The following figure 2 shows how CBR works when diagnosing breast cancer.

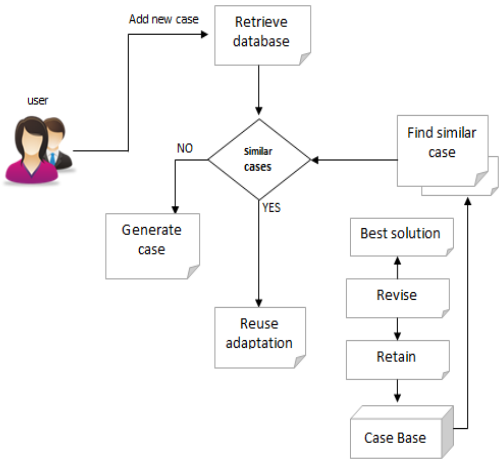

Fig. 2. Breast cancer framework using CBR

The patient enters the symptoms he suffers in the application and this application returns the preceding cases and sees if there is a similar case or not, if there are no similar cases then the answer is NO (does not have breast cancer) and adds the case to the new cases in the application, and if the answer is YES (have breast cancer) then the old case is similar to the new case.

4. Analysis of Result

If you'd like to explore the linear regression more, dataquest offers an excellent course on its use and application! We used to scikit-learn to apply the error metrics in this article, so you can read the docs to get a better look at how to use them! The shown similarity table and its corresponding column plot (bar chart) demonstrate the closeness of each of the base cases (labeled C1, C2, C3, C4, C5, C6, C7, C8, C9, and C10) to a new case. There is a bar assigned to each base case (based on the coloring illustrated in the figure legend). A new case is considered matching with the base case having the smallest similarity function, for example C8, and this base case is then chosen. It is important to say that the distance between a pair of cases (a base case and a new case) can be measured in more than one algorithm or method. Here, five distance algorithms were used, which are: Euclidian, Manhattan, Canberra, squared chord, and squared chi-squared. An algorithm can show better performance than others. The accuracy of a diagnosis or detection system is a percentage number ranging from 0% (worst case) to 100% (best case). It is the proportion of correctly detected cases out of the totally detected ones.

$$Account\ Accuracy = \frac{Correct\ Diagnosed}{Total\ Testing\ Cases} \times 100 \qquad (1)$$

The equation (1) gives the mathematical formula for the accuracy. In this paper, the total testing cases is fixed as 10. The number of correct diagnosed cases in the first similarity function (Manhattan) is 5. Therefore, its accuracy is 50%. This is also the accuracy found for the second function (Euclidian). For the third function (Canberra), the correct diagnosis value drops to 0.9, giving a much smaller accuracy of 9%. The fourth function (Squared chord) has a better accuracy of 25%. Finally, the last function (Squared chi-squared) has an accuracy of 13%. The equations below and the table 2 following them give a summary of accuracy values for all the five similarity functions of in this paper and how they were calculated numerically.

Table 2
Accuracy of functions

Function	Accuracy
Manhattan	0.5%
Euclidian	0.5%
Canberra	0.09%
Squared Chord	0.025%
Squared chi-squared	0.013%

The accuracy of a detection or diagnosis process is a measure of its ability to correctly diagnose a given case as depicted in figure 3 and table 3. There are two types of mistaken diagnosis; either a positive case is diagnosed as negative, or a negative case is diagnosed as positive. Both situations are undesirable. The accuracy is calculated as the fraction of test cases (whose true medical conditions are known) that are correctly diagnosed. This means dividing the number of correctly diagnosed cases by the total number of diagnosed cases (correctly diagnosed and incorrectly diagnosed). Finally, this fraction is expressed as a percentage through multiplying by 100%. There is a number of algorithms for measuring distances between points or objects.

One is the Euclidean distance is gives the shortest, straight-line, separation between any two points.

The Manhattan distance is somehow longer than the Euclidean distance, where one is restricted to travel segments at right angles (for example, purely horizontal and purely vertical).

The Canberra distance is not very different from the Manhattan distance. The difference is that the magnitude of the difference between the coordinates of the two points is divided by the sum of the magnitude values of the coordinates before the summation.

Other distance measures are the squared chord distance, and the chi-square distance.

There are formulas for the similarity calculations based on the distance algorithm. A computer program calculates the value of the similarity function to the chosen base case after a user enters the symptoms setting for a new case to be diagnosed as showed in figure 4.

Table 3
Cases accuracy of similarity functions

Function	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	Chosen
Manhattan	3	3	5	3	3	2	2	2	5	3	8
Euclidean	0.6	0.6	0.933	0.6	0.6	0.33	0.33	0.33	0.933	0.6	8
S-Chord	1	1	1.434	1	1	0.343	0.343	0.343	1.343	1	8
S-chi-s	1.8	1.8	2.466	1.8	1.8	0.66	0.66	0.66	2.466	1.8	8

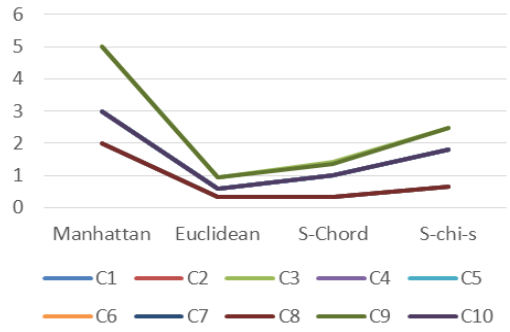

Fig. 3. Cases accuracy rate

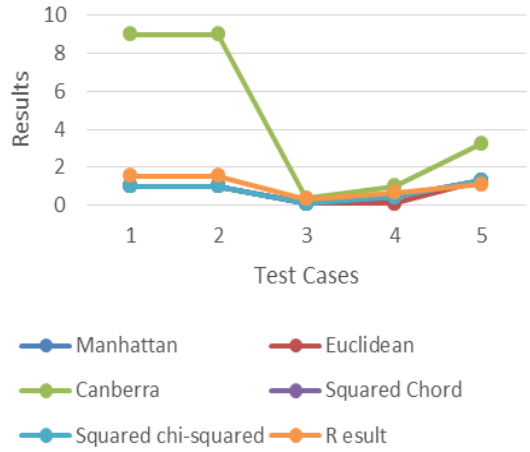

Fig. 4. Similarity functions accuracy rate

$$\text{RMSE}(\text{X1}.\text{X2}) = \frac{\sqrt{\sum_{i=0}^{n} (x1-x2)*2}}{n} \quad (2)$$

The equation (2) of root-mean-square error (RMSE) measure of deviation is calculated from the similarity function table of each distance algorithm as follows: First, the difference between the similarity of each base case and the chosen case is computed, this gives an array of numbers with a size (n). Then, these differences are squared; this is important to eliminate misleading cancellation of deviations having opposite signs. Then, the average (the mean) of the squared values is obtained by dividing their sum over their number (n). Finally, the square root of this average is taken, and the results is taken as the RMSE. The RMSE for each distance algorithm as calculated using the above steps is listed in the table 4 below.

Table 4
Error rate breast cancer using RMES

Function	Error rate
Manhattan	1.517
Euclidean	1.517
Canberra	0.3304
Squared Chord	0.6448
Squared chi-squared	1.1411

The table 5 below described the numerical calculation of the RMSE for each similarity function (each distance algorithm). The top row in each function represents input similarity values taken. The second row is representing the differences with respect to the new case (having the smallest similarity) after being squared. Finally, the square root of the average of the numbers in the second row gives the RMSE. This is done for each of the five functions.

The figure 5 below described the numerical calculation of the RMSE for each similarity function (each distance algorithm). The top row in each function represents input similarity values taken. The second row represents the differences with respect to the new case (having the smallest similarity) after being squared. Finally, the square root of the average of the numbers in the second row gives the RMSE. This is done for each of the five functions.

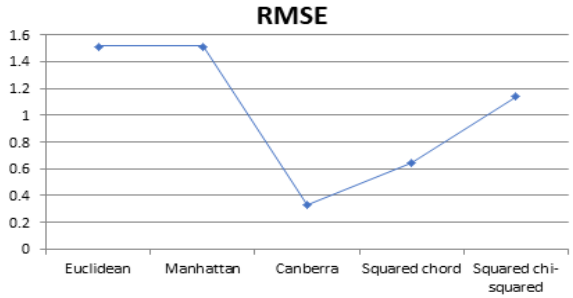

Fig. 5. Error rate of similarity functions using RMSE formula

5. Conclusion

The result of the testing does not achieve 100%; but we achieved a high percentage of accuracy. CBR approaches help mitigate the lack of young medical personnel expertise. The inexperience staff need the guidance from the experience staff to improve their diagnostic handling skills. We have a lot of contributions that we got. The recommendations that should be done in order to improve the application as follow:

1. This application can be implemented for next version inside the mobile application due to technology

Table 5
Error rate of similarities functions

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	New case	Function
	3	3	5	3	3	2	2	2	5	3	8	Euclidean
(x1-x2)^2	1	1	9	1	1	0	0	0	9	1		
	3	3	5	3	3	2	2	2	5	3	8	Manhattan
(x1-x2)^2	1	1	9	1	1	0	0	0	9	1		
	0.6	0.6	0.933	0.6	0.6	0.33	0.33	0.33	0.93	0.6	8	Canberra
(x1-x2)^2	0.07	0.072	0.363	0.072	0.072	0	0	0	0.36	0.07		
	1	1	1.343	1	1	0.34	0.343	0.343	1.34	1	8	Squared chord
(x1-x2)^2	0.43	0.431	1	0.431	0.431	0	0	0	1	0.43		
	1.8	1.8	2.466	1.8	1.8	0.66	0.66	0.66	2.46	1.8	8	Squared chi-squared
(x1-x2)^2	1.29	1.299	3.261	1.299	1.299	0	0	0	3.26	1.29		

development nowadays.

2. Meet the medical expertise on breast cancer to find out the most important attributes they used in breast cancer diagnosis.
3. Develop the application to include other diseases of cancer.

References

[1] Nadir K Salih, Tianyi Zang. Variable service process for SaaS Application. Research Journal of Applied Sciences, Engineering and Technology. vol. 4, no. 22, pp. 4787-4790, 2012.

[2] Nadir K Salih, Tianyi Zang, Mingrui Sun. Multi-database in healthcare network. International Journal of Computer Science Issues, vol. 8, no. 6, pp. 210-214, 2011.

[3] Nadir K Salih, Tianyi Zang, G.K. Viju, A Mohamed. Autonomic management for multi-agent system. IJCSI, vol. 8, no. 5, pp. 338-341, 2011.

[4] Nadir K Salih, Tianyi Zang. Need of Autonomic Management SaaS Application. International Journal of Computer Science Issues, 2016.

[5] Nadir K Salih, Tianyi Zang. Survey and comparison for Open and closed sources in cloud Computing. International Journal of Computer Science Issues, vol. 9, no. 3, 2012, pp. 118-123.

[6] Eman M-Fageer, Nadir K. Salih. Self-configuring Booking SaaS Application. Red Sea University Journal of Basic and Applied Science, vol. 2 Special Issue (3), 2017.

[7] Amin, Fatima M H, Nadir K. Salih. New Model to Achieve Software Quality Assurance in E-Learning Application. International Journal of Computer Science Issues, pp. 65-69, May 2017.

[8] Eshtiag A Abd Elrhman, Nadir K Salih. Modeling Variation in SaaS Application. International Journal of Computer Science Issues, vol. 15, no. 3, pp. 22-30.2018.

[9] Salih N. K, H. Elbashier, Zang T, Eshtiag A. Abd Elrhman. Self-Diagnosis of Diabetes Using CBR Algorithm. Journal of Computer Science & Systems Biology. Vol. 11, no. 3, pp. 235-239, 2018.

[10] Amin, Fatima M H, Nadir K. Salih. Implementing the System, Instructor and Student Model to Achieve Required Software Quality Assurance. Research Journal of Applied Sciences, Engineering and Technology, pp. 30-42, 2019.

[11] Nadir K. Salih, Abdel-hafiz A. Khoudour, Mawahib S. Adam, Samar M. Hassen, "Autonomic Computing Architecture by Self-defined URI" International Journal of Computer Trends and Technology, 2020.

[12] Nadir K Salih, Tianyi Zang. Variable service process by feature meta-model for SaaS Application. IEEE International Conference in Green and Ubiquitous Technology, pp. 102 –105, 2012.

[13] Nadir K Salih, Tianyi Zang. Autonomic and cloud computing: Management Services for Healthcare. IEEE International Symposium on Industrial Electronics and Applications (ISIEA 2012).

[14] Nadir K Salih, Tianyi Zang. Modeling and Self-Configuring SaaS Application. International conference on software engineering research and practice (SERP14), July 21-24, Las Vegas, USA, 2014.

[15] Nadir K Salih, Tianyi Zang. Autonomic Management for Applicability and Peformence in SaaS Model. International conference on parallel and distributed processing techniques and applications (PDPTA'14), July 21-24, Las Vegas, USA, 2014.

[16] Nadir K Salih, Tianyi Zang. Self-management SaaS Application by CBR Algorithm. International conference on parallel and distributed processing techniques and applications (PDPTA'17), July 21-24 Las Vegas, USA, 2017.

[17] Nadir K. Salih, Tianyi Zang. Implementation of Autonomic Management SaaS System. conference on software engineering research and practice (SERP14), July 21-24, Las Vegas, USA 2017.

[18] GK Viju, Nadir K Salih, Tianyi Zang. A novel approach to iris recognition for personal authentication. International Conference of Computer Applications and Industrial Electronics (ICCAIE), 2011, pp. 350-354.

[19] G. K. Viju, Nadir K. Salih. A secure multicast protocol for ownership rights. International Conference of Computing and Information Technology (ICCIT), 2012, pp. 788-793.

[20] Sheima S. El-hwaij, Nadir K. Salih. Autonomic management by self-optimization for WEINMANN. IEEE International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE), 2017.