Advances in Engineering Research, volume 210
Conference on Broad Exposure to Science and Technology 2021 (BEST 2021)

A Case-Based Reasoning for Detection Coronavirus (Covid-19) Using Cosine Similarity

Murien Nugraheni[1*], Widodo, Irma Permata Sari[1]

[1]*Information System and Technology, Faculty of Engineering, State University of Jakarta, Indonesia*
[*]*Corresponding author. Email: muriennugraheni@unj.ac.id*

ABSTRACT
Case-based reasoning is a new approach that can be used to diagnose disease in addition to using expert systems or other approaches, which are part of artificial intelligence. Case-based reasoning can diagnose diseases based on visible or perceived clinical symptoms. This study tries to build case-based reasoning for early detection of COVID-19 by looking at the characteristics of clinical symptoms seen in a person using the Cosine Similarity method. Cosine similarity is a method to find level of similarity between two cases. The detection process is carried out by entering a new case containing symptoms into the system, then system will perform a similarity calculation process between the old case and the new case. The results show case-based reasoning for early detection of COVID-19 using the Cosine Similarity method can detect a similarity level of 80%.

Keywords: *Coronavirus Detection, Artificial Intelligence, Early Detection, Cosine Similarity Method*

1. INTRODUCTION

Coronavirus disease 2019 (COVID-2019) is a disease that is currently endemic around the world and first reported on December 31, 2019 in Wuhan, Hubei Province, China, which reported the first cases of pneumonia with no known cause [1]. The COVID-19 pandemic caused an economic shock, closure of cities and countries with aim of limiting the spread of the virus [2]. Coronavirus is a unsegmented single-strain RNA virus and encapsulated [3]. Symptoms of fever, cough, runny nose, and shortness of breath are some of the symptoms of COVID-19 that are often felt [4]. COVID-19 spread via respiratory droplets produced when people sneeze or cough [5]. The incubation period ranges from 5 to 14 days [4]. So proper and frequent hand washing is so important [5]. Some of the diseases caused by coronavirus include kidney failure, pneumonia, acute respiratory failure, and death [4]. The infiltrative pneumonia in the lungs is shown by x-ray result with the perceives symptoms are difficulty breathing [4]. The steps to detect COVID-19, one of which is using x-ray. X-ray have very short wavelengths and are emission of electromagnetic waves similar to light wavelengths, heat, radio, and ultraviolet [6]. The purpose of this study was to determine the level of similarity for early detection of coronavirus using the cosine similarity method in Case-based Reasoning (CBR) and is expected to detect coronavirus disease more accurately. So our main contribution is to be able to detect coronavirus disease in this pandemic by applying CBR by calculating the level of similarity using the cosine similarity method.

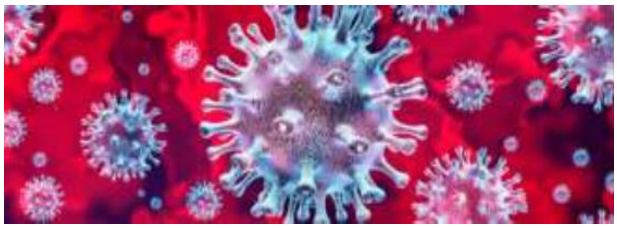


Figure 1. Corona virus image

Table 1 present the statistics of number of total cases, total deaths, total recovered, and active cases worldwide in the top 30 countries sorted in descending order by the number of total cases as of August 30, 2020. [7].

Artificial intelligence is how to build computers or design machines to similar human abilities. One of the benefits of artificial intelligence is the infection re-sults see on CT scan and x-ray in improve work effec-tiveness and efficiency[8]. Some example of artificial intelligence

Table 1. Statistics of total cases, total deaths, total recovered, and active cases

No	Country	Total Cases	Total Deaths	Total Recovered	Active Cases
1.	USA	6.139.466	186.857	3.408.908	2.543.701
2.	Brazil	3.846.965	120.498	3.006.812	719.655
3.	India	3.546.705	63.690	2.714.995	768.020
4.	Russia	990.326	17.093	806.982	166.251
5.	Peru	639.435	28.607	446.675	164.153
6.	South Africa	622.551	13.981	536.694	71.876
7.	Colombia	599.914	19.064	440.574	140.276
8.	Mexico	591.712	63.819	409.127	118.766
9.	Spain	455.621	29.011	N/A	N/A
10.	Chile	408.009	11.181	381.183	15.645
11.	Argentina	401.239	8.353	287.220	105.666
12.	Iran	371.816	21.359	319.847	30.610
13.	UK	332.752	41.498	N/A	N/A
14.	Saudi Arabia	313.911	3.840	288.441	21.630
15.	Bangladesh	308.925	4.206	198.863	105.856
16.	Pakistan	295.636	6.288	280.547	8.801
17.	France	272.530	30.602	86.177	155.751
18.	Turkey	267.064	6.284	242.812	17.968
19.	Italy	266.853	35.473	208.224	23.156
20.	Germany	242.825	9.363	217.484	15.978
21.	Iraq	227.446	6.891	169.020	51.535
22.	Philippines	213.131	3.419	135.101	74.611
23.	Indonesia	169.195	7.261	122.802	39.132
24.	Canada	127.673	9.113	113.501	5.059
25.	Ukraine	119.074	2.527	56.734	59.813
26.	Qatar	118.407	196	115.251	2.960
27.	Bolivia	115.354	4.938	56.749	53.667
28.	Israel	113.623	906	92.587	20.130
29.	Ecuador	112.906	6.537	95.429	10.940
30.	Kazakhstan	105.684	1.523	96.135	8.026

is a deep learning features, expert systems, neural networking, case-based reasoning and etc. Deep learning features can be used tofind out one of the symptoms of COVID-19 utilizing x-ray[5]. Deep learn-ing algorithms are widely used because it has higher advantages compared to other learning algorithms [9].

Beside using deep learning model can be used CBR, expert system, and artificial neural network. The expert system is expected to be able to help doctors and medical specializations to diagnose the disease from the symptoms felt or seen by COVID-19 patients, so that a faster and more accurate diagnosis is expected [10]. The SL5 is an object language by utilizing an expert system to diagnose patients exposed to coronavirus and is given directions or instruction that must be followed [11]. The one of the neural network model is CoroNet[12]. CBR models make use of previous knowledge of patient cases or real problem situation to solve new cases [13]. One method that can be used to find the closeness or similarity of the new case to the old case in CBR is Cosine Similar-

ity. This method correlates two continuous variable objects that have a linear relationship between the attributes of an object.

2. METHODS

2.1. Case-based Reasoning

Computer reasoning by utilizing old cases to be used as solutions to new cases is called CBR. In addition, CBR can also be started from a small amount of knowledge because case-based reasoning knowledge can gradually increase as a case is added. CBR emphasizes problem solving based on the knowledge of previous cases, when there is a new case, it is stored in a knowledge base so that the system perform learning and knowledge possessed by the system increase.

CBR can be represented as a process cycle which is divided into four sub processes [14]:

- Retrieve which is looking for old cases that are most similar to new case.
- Reuse is reusing the most similar cases to get a solution for a new case.
- Revise is making adjustments to the solution of previous cases so that they can be used as solutions for new cases.
- Retain is to use a new solution of a new case, then the new case is updated on a case basis.

2.2. Cosine Similarity

Cosine Similarity is one of the similarity methods used to calculate the similarity level of two objects. Cosine similarity can be formulated as follows: [15]

$$\text{Cos}(x,y) = \frac{x \cdot y}{\| x \| \ \| y \|} \quad (1)$$

with $x.y = \sum_{k=1}^{n} x_k y_k$

$\| x \|$ is length of the vector x, $\| x \| = \sqrt{\sum_{k=1}^{n} x_k^2}$

$\| y \|$ is length of the vector y, $\| y \| = \sqrt{\sum_{k=1}^{n} y_k^2}$

x is value from of old case with $x = x_1, x_2, x_3, \ldots x_n$

y is value from of new case with $y = y_1, y_2, y_3, \ldots y_n$

The higher similarity value, the higher level of similarity to old case.

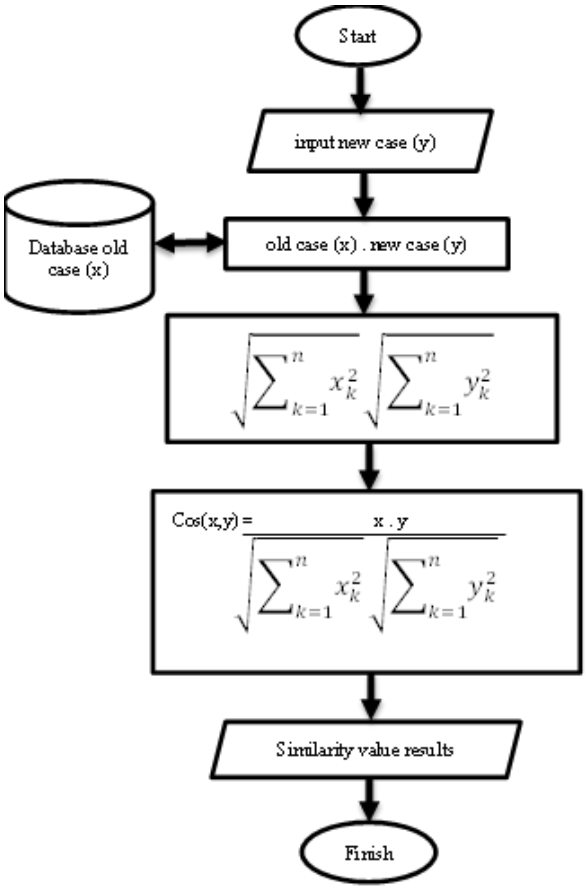


Figure 2. Flowchart cosine similarity

From Figure 2, steps in Cosine Similarity method can be explained. First, a new case (y) be input into system then a multiplication calculation is performed from the old case (x) and new case (y), where old case (x) is obtained from case base stored in database. Then multiplication calculation between norm x (||x||) dan norm y (||y||). The last, calculation Cosine Similarity formula which produce output similarity level fromf old case (x) and new case (y).

2.3. Coronavirus Disease 2019 (COVID-19)

COVID-19 is a new disease that has become a pandemic. COVID-19 disease must be watched out for because the transmission is so fast, it has a mortality rate that needs attention, and there is no definite therapy[16]. In Table 2 several groups of COVID-19 patients are also shown, namely “PatientUnder Surveillance or Suspect” and “People in Monitoring”. Some of the symptoms of COVID-19 can be seen in Table 2.

Table 2. Symptoms patient under surveillance and people in monitoring

Findings	Patient Under Surveillance/ Suspect				People in Monitoring	
	Criterion 1	Criterion 2	Criterion 3	Criterion 4		
1. Fever/ history of fever	Yes	Yes	Yes	-	For either of these two points, for no other reason is clear	Yes
2. Cough/ cold/ sore throat/ hard to breathe (one of choice)	Yes	-	Yes	-		
3. Travel to a pandemic country area (the last 14 days)	Yes	-	-	Yes	Yes	
4. Contact with COVID-19 confirmation case	-	Yes	Yes	-		Yes
5. Pneumonia/ ISPA without cause	-	-	-	Yes		

3. RESULTS AND DISCUSSION

3.1. Case Representations

In this research there were 9 symptoms can be seen in Table 3 and 20 data samples, where later similarity calculations performed for new cases, as many as 10 cases. Table 4. Shows the relationship between symptoms, classes, and cases. In the table there are 21 cases, namely CS01 to CS021, with 9 symptoms, namely G001 to G009, each of which has a value of "no" (0) and "yes" (1). In addition, there are also classes categorized by CL1 patient under surveillance, CL2 is people in monitoring, and CL3 people without symptoms.

Table 3. Symptoms data

Symptoms Code	Symptoms Name
G001	Fever
G002	Cough
G003	Cold
G004	Sore throat
G005	Hard to breathe
G006	Travel to an infected country
G007	Contact with covid-19 confirmation case
G008	Pneumonia
G009	ISPA

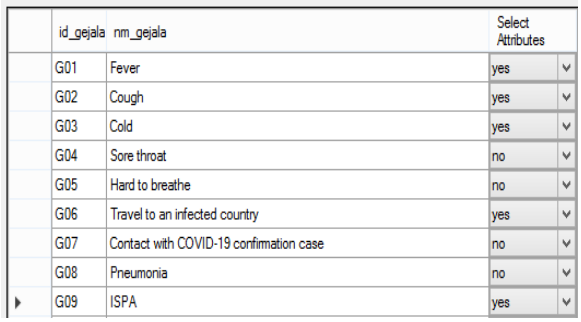
In CBR, the old case be stored in the base case which be used to find the similarity value for the new case. If there are new cases that have not been saved on the case base, then the new cases be stored in the case base to be reused as a knowledge base.

3.2. Acquiring New Cases Stage

The user inputs the symptoms that are seen or felt. Then the calculation is carried out to find the similarity value between the new case and the old case stored on a case base, if there is a similarity then the solution from the old case is given to the user. To find the closeness between new cases and old cases stored in the case base using the Cosine Similarity as a similarity method.

3.3. Results

Enter any visible symptoms or symptoms felt in system. Where a value of 1 indicates "yes" that is symptoms is felt. While a value of 0 indicates "No" that is symptoms is not felt. Figure 3 is a display of symptoms that exist in the system according to the list of symptoms in Table 3. There are 9 symptoms that can be selected by user according to the symptoms feel.



	id_gejala	nm_gejala	Select Attributes
	G01	Fever	yes
	G02	Cough	yes
	G03	Cold	yes
	G04	Sore throat	no
	G05	Hard to breathe	no
	G06	Travel to an infected country	yes
	G07	Contact with COVID-19 confirmation case	no
	G08	Pneumonia	no
▶	G09	ISPA	yes

Figure 3. Form new cases diagnosis

The test result for 10 cases can be seen in Table 5. Of the 10 test cases, there were 8 cases with a similarity above 0,75 and 2 cases with a similarity less than 0,75 so cases with a similarity of more than 0,75 were 80 % and

Table 4. The relationship between symptoms, classes, and cases.

Cases	G001	G002	G003	G004	G005	G006	G007	G008	G009	Class
CS01	1	1	0	0	0	1	0	0	0	CL1
CS02	1	0	1	0	0	1	0	0	0	CL1
CS03	1	0	0	1	0	1	0	0	0	CL1
CS04	1	0	0	0	1	1	0	0	0	CL1
CS05	1	0	0	0	0	0	1	0	0	CL1
CS06	1	1	0	0	0	0	1	0	0	CL1
CS07	1	0	1	0	0	0	1	0	0	CL1
CS08	1	0	0	1	0	0	1	0	0	CL1
CS09	1	0	0	0	1	0	1	0	0	CL1
CS010	0	0	0	0	0	1	0	1	0	CL1
CS011	0	0	0	0	0	1	0	0	1	CL1
CS012	1	0	0	0	0	1	0	0	0	CL2
CS013	0	1	0	0	0	1	0	0	0	CL2
CS014	0	0	1	0	0	1	0	0	0	CL2
CS015	0	0	0	1	0	1	0	0	0	CL2
CS016	0	0	0	0	1	1	0	0	0	CL2
CS017	0	1	0	0	0	0	1	0	0	CL2
CS018	0	0	1	0	0	0	1	0	0	CL2
CS019	0	0	0	1	0	0	1	0	0	CL2
CS020	0	0	0	0	1	0	1	0	0	CL2
CS021	0	0	0	0	0	0	0	0	0	CL3

Table 5. The similarity value of test cases

No	New Cases	Class	The Closest Old Case	Similarity Value
1.	CS031	People in monitoring	CS018	0,8165
2.	CS032	Patient under surveillance	CS006	0,7746
3.	CS033	People in monitoring	CS017	0,8165
4.	CS034	Patient under surveillance	CS002	0,8165
5.	CS035	Patient under surveillance	CS001	0,86603
6.	CS036	Patient under surveillance	CS006	0,70711
7.	CS037	People in monitoring	CS017	0,86603
8.	CS038	Patient under surveillance	CS003	0,8165
9.	CS039	Patient under surveillance	CS017	0,70711
10.	CS040	Patient under surveillance	CS002	0,7746

cases with a similarity less than 0,75 were 20 %. The similarity value for each new case was obtained by showing the old case which was the closest seen from the inputted symptoms. Although there are several that have the same similarity value, the results of the old cases stored in the case base have different results, because the system looks for the symptoms stored in the old cases that are closest to those of the test case. The higher similarity value, the higher level of similarity to old case. From the old case can be used for solutions to new cases that are input to

system.

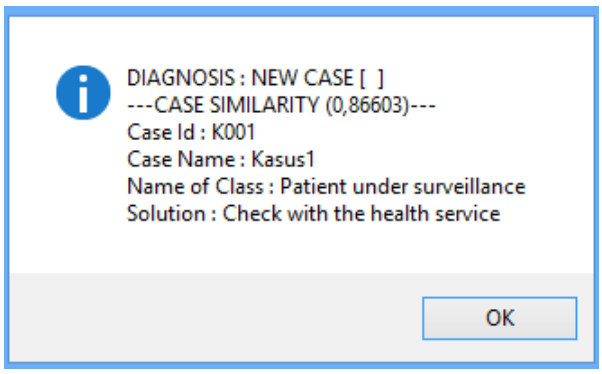


Figure 4. New cases diagnosis result

4. CONCLUSION

The results of research Case-Based Reasoning for Detection Coronavirus (COVID-19) using Cosine Similarity, it can be concluded that:

- Detection of Coronavirus cases used CBR.
- Based on the results of testing on 10 cases, there were 8 cases with a similarity value above 0,75 and 2 cases with a value below 0,75. So get cases with a similarity value of more than 0.75 by 80 % and cas-es with a similarity value of less than 0.75 by 20 %.

It is hoped that in the development of Case-Based Reasoning for Detection Coronavirus (COVID-19) using Cosine Similarity, research can be carried out on other similarity methods which can produce a better similarity value. And can use indexing methods to make easier to find similarity on a very large number of cases.

REFERENCES

[1] Pokja Infeksi Saluran Reproduksi Perkumpulan Obstetri Ginekologi Indonesia (POKJA ISR PP POGI) 2020 Rekomendasi Penanganan Infeksi Virus Corona (Covid-19) pada Maternal (Hamil, Bersalin, dan Nifas) Pokja Infeksi Saluran Reproduksi : Perkumpulan Obstetri dan Ginekologi Indonesia Jakarta

[2] Rashwan A R M S and Alhelou E M S 2020 The Effect of the Covid-19 Pandemic on the Company's Suistainability under International Accounting Standards, International Journal of Academic Accounting, Finance & Management Research 4(7) 34-51

[3] Perhimpunan Dokter Paru Indonesia (PDPI) 2020 Pneumonia Covid-19 : Diagnosis dan Penatalaksanaan di Indonesia Perhimpunan Dokter Paru Indonesia Jakarta

[4] Kemenkes RI 2020 Pedoman Pencegahan dan Pengendalian Coronavirus Disease (Covid-19) (Jakarta : Kementerian Kesehatan RI Direktorat Jenderal Pencegahan dan Pengendalian Penyakit (P2P))

[5] Salman F M, Naser S S A, Alajrami E, Nasser B S A and Ashqar B A M 2020 COVID-19 Detection using Artificial Intelligence International Journal of Academic Engineering Research 4(3) 18-25

[6] Suryan D 2018 Analisis Gelombang Elektromagnetik (Sinar-X) dalam Kesehatan Perspektif Al-Qur'an dan Sains Prosiding Seminar Nasional Pendidikan Fisika FTTK UNSIQ 1(1) 115-120

[7] World Health Organization (WHO)

[8] Shi F, Wang J, Shi J, Wu Z, Wang Q, Tang Z, He K, Shi Y, and Shen D 2020, Review of Artificial Intelligence Techniques in Imaging Data Acquisition Segmentation and Diagnosis for COVID-19 IEEE Reviews in Biomedical Engineering pp 1-13

[9] Lalmuanawa S, Hussain J, and Chhakchhuak L 2020 Applications of Machine Learning and Artificial Intelligence for COVID-19 (SARS-CoV-2) Pandemic : A review Elsevier Chaos Solitons and Fractals 139 pp 1-6

[10] Salman F M and Naser S S A 2020 Expert System for COVID-19 Diagnosis International Journal of Academic Engineering Research 4(3) 1-13

[11] Almadhoun H R and Naser S S A 2020 An Expert System for Diagnosing Coronavirus (COVID-19) using SL5 International Journal of Academic Engineering Research 4(4) 1-9

[12] Khan A I, Shah JL, and Bhat M M 2020 CoroNet : A Deep Neural Network for Detection and Diagnosis of COVID-19 from Chest X-Ray Images Elsevier Computer Methods and Programs in Biomedicine 196 pp 1-9

[13] Oyelade O N and Ezugwu A E 2020 A Case-based Reasoning Framework for Early Detection and Diagnosis of Novel Coronavirus Elsevier Informatics in Medicine Unlocked 20 pp 1-22

[14] Aamodt A and Plaza E 1994 Case-Based Reasoning: Foundational Issues Methodological Variations and System Approches AICom-Artificial Intelligence Communications IOS Press 7(1) 39-59

[15] Tan P N, Steinbach M, Kumar V 2006 Data Mining (Pearson Education, Inc)

[16] Susilo A, Rumende CM, Pitoyo C W, Santoso W D, Yulianti M, Herikurniawan Sinto R, Singh G, Nainggolan L, Nelwan E J, Chen L K, Widhani A, Wijaya E, Wicaksana B, Maksum M, Annisa F, Jasirwan C O, and Yunihastuti E 2020 Coronavirus Disease 2019 : Tinjauuan Literatur Terkini Junal Penyakit Dalam Indonesia (Indonesian Journal of Internal Medicine) 1(2) 45-67.