Date of publication JUN-28, 2021, date of current version JUL-28, 2020.
www.computingonline.net / computing@computingonline.net

Print ISSN 1727-6209
Online ISSN 2312-5381
DOI 10.47839/ijc.20.2.2174

The Efficient Distance Weighted Case Base Rule (DW-CBR) for Early Childhood Diseases Diagnosis

INDAH WERDININGSIH[1], RIMULJO HENDRADI[1], PURBANDINI[1], BARRY NUQOBA[1], ELLY ANNA[2]

[1]Information System Study Program, Department of Mathematics
Universitas Airlangga, Mulyorejo Street, 60115, Surabaya, Indonesia
(e-mail: indah-w@fst.unair.ac.id, rimuljohendradi@fst.unair.ac.id, purbandini@fst.unair.ac.id, barrynuqoba@fst.unair.ac.id)
[2]Statistics Study Program, Department of Mathematics
Universitas Airlangga, Mulyorejo Street, 60115, Surabaya, Indonesia
(e-mail: elly-a@fst.unair.ac.id)

Corresponding author: Rimuljo Hendradi (e-mail: rimuljohendradi@fst.unair.ac.id).

ABSTRACT Children from newborns to six years old are more susceptible to diseases. A common methodology to diagnose childhood diseases is by using a reasoning technique. Reasoning techniques is one of a reliable method for expert systems. Reasoning techniques using the correct case of results have provided enormous support for predicting the diagnosis and treatment of diseases. This paper focuses on the main technical characteristics of two common reasoning techniques, namely; rule-based reasoning and case-based reasoning. This paper describes a comparative analysis of rule-based and case-based reasoning techniques using several commonly used similarity measures and a study on its performance for classification tasks. Moreover, this study proposes a new case-based reasoning approach using an alternative similarity measure, called Distance-Weighted Case Base Reasoning (DW-CBR). The proposed method aims to improve classification performance. The main result of this study shows that case-based reasoning is a more powerful methodology regarding the issues of maintenance and knowledge representations over the rule-based system and reveals that DWCBR has the best accuracy, which is 92%.

KEYWORDS Expert System; Childhood; Reasoning; Similarity; Nearest Neighbor; Rule Based.

I. INTRODUCTION

Young children are very susceptible to disease [1]. More than 400 children in Indonesia die from illness per day, with the most common conditions being pneumonia and diarrhea [2]. Research also shows that early childhood mortality rates are around 65 per 1,000 births [3]. Children's health in Indonesia, therefore, is far from the standard set by the World Health Organization (WHO) in 1946, which stated that all children have the right to good health and grow to be world citizens [4].

Clinical diagnosis is performed based on the doctor's advice and opinions. Doctors usually recommend taking several tests to diagnose a disease, although most of the tests are unnecessary. Hidden patterns in medical databases are relatively uncharted in medical diagnosis despite their great potential. One problem in medical databases is that they are high in volumes, so chances of errors in diagnosis, treatment, and classification are also high.

Knowledge management and database management systems are already widely used in many decision support systems (DSS) [5]. There is an actual need to create knowledge management systems that accumulate experience for future decision support in certain situations. If not, we can face the lack of specific information and make some wrong decisions, sometimes vital for someone. The reasoning technique comes as a solution to increase the efficiency of decision making using the accumulated data.

Machine learning has a powerful adaptive learning ability from which it derives its strong interpolative capability. That ability makes it very suitable for prediction, especially in instances of noisy data or incomplete data, which many other alternative prediction models are not able to handle [6]. However, machine learning has a fragile explanation mechanism, which makes it challenging to understand the reasoning behind its conclusions [7]. That weakness is a significant limitation, particularly in the diagnosing and treatment of diseases where it is essential to have a strong rationale for making decisions.

The reasoning technique is an excellent tool for making predictions, and the medical field has used it extensively. Reasoning techniques using the correct case of results have provided exceptional support for predicting the diagnosing and treatment of diseases.

Case-based reasoning can be particularly useful in areas where traditional rule-based reasoning is relatively weak, such as knowledge acquisition, machine learning, and reasoning with incomplete information [8]. CBR is an approach using previous experience cases to solve a new case [9, 10]. CBR is a machine learning paradigm that models the human reasoning process. CBR systems have a robust explanation mechanism because of the existence of sufficiently similar previous cases that provides a good rationale for new solutions obtained.

Case-based reasoning (CBR) is one of the most commonly used reasoning techniques. CBR is the more efficient, robust and less cost [11]. Typically, a complete schematic of building CBR systems includes four steps: case retrieval, case reuse, case revision, and case retention, of which in particular case retrieval needs a specific concern about the anticipated outcome of system design [9, 12].

Technically, the similarity measurement plays a significant role in the process of case retrieval [13]. Currently, several similarity similar measures have been used in CBR System [14]. Similarity measures are intrinsically related to the used case representation formalism in the CBR system. Computation of the similarity measures is commonly by aggregating the attribute differences between the two cases. If similarity measures do not capture the actual differences between cases, the retrieval step, and the whole CBR will have a poor performance. Therefore, the selection of an appropriate similarity measure in the retrieval step is a crucial point in CBR systems [15].

Nearest Neighbor (NN) is one of the most commonly used similarity measures. The NN classifies new objects based on their nearest neighbors, with its simple and relatively high speed of convergence [16]. The NN idea is that an unknown instance will be similar to others close to it in terms of characteristics. Although this idea is simple, it requires high computation because the number of operations will increase as the dataset's size increases, both in terms of attributes and instances. When handling large problems, the large amount of time needed to calculate results makes NN almost unusable [17].

The selection of the neighborhood that exists in CBR is a sensitive problem. The distance of the nearest neighbor to the query determines the radius of the local region. Different radius yields different conditional class probabilities. If the length of the neighborhood is minimal, the local estimate tends to be very poor, owing to the data sparseness and the noisy, ambiguous, or mislabeled points [18].

In this study, we proposed a modified CBR using an alternative similarity measure, which is then called Distance-Weighted Case Base Reasoning (DW-CBR). The experiments compared the DW-CBR against some other measurements that had been used before, with an excellent performance in tests done in prior studies were carried out. DW-CBR aims to improve classification performance. DW-CBR used the basic idea of close neighbor weighting, according to their distance from the query [19]. We propose DW-CBR using the dual distance-weighted function. In this new rule, we employ the dual distance-weights of nearest neighbors to determine the class of the query by majority weighted voting. To summarize, the significant contribution of this paper includes:

- An alternative similarity measure proposed in CBR, which is called Distance-Weighted Case Base Reasoning (DW-CBR)
- A new weighting mechanism employed by using dual distance-weights of nearest neighbors to determine the class of the query by majority weighted voting
- Developing an efficient reasoning technique for early childhood disease expert systems.

In what follows, the reasoning techniques and details of the proposed approach are presented, continued with the description of the research results and discussion, which is then followed by the inferred conclusions.

II. REASONING TECHNIQUES IN EXPERT SYSTEM

Inference, reasoning, and learning abilities are the main features of an expert system. In this section, we focus the discussion on the main characteristics of the two reasoning techniques commonly used in the development of early childhood expert systems, i.e., rule-based reasoning and case-based reasoning (CBR).

A. RULE BASED REASONING

Rule-based reasoning used in this study was forward chaining because the forward chaining more accurately than backward chaining [20]. Forward chaining is part of rule-based reasoning, which are integral parts of the development of expert-system inference machines [21]. Forward chaining usually starts with a set of facts and adding a statement every time there is a new fact.

Forward chaining is an antecedent reasoning process that drives new facts based on a set of rules or an initial set of data. The antecedent or data-driven rules are used when a user's response to a question, or in a less common situation, a conclusion derived from another rule, triggers a file system's forward-chained rules. These rules' premises have a single condition; only if one rule concludes with absolute certainty that a premise is true, then repeated forward chaining is permitted. New data is generated by a simple and straightforward application of the rule firing [22]. Forward chaining is a swift inference procedure and is best used in monitoring and in a diagnostic system to identify and respond quickly to real problems [21].

In Forward Chaining, if the premise (if) is correct, the conclusion will also be correct. The steps to search in a Forward chaining technique are:

Step 1: Asking questions to a user
Step 2: Receiving input from a user as known facts in short-term memory that are stored in each variable (the question being asked)
Step 3: Checking the rule based on the facts stored in the short-term memory by using forward chaining method
Step 4: If rules are found, then conclusions are accommodated in short-term memory; if there are new facts, then step one through step four is repeated. If the rule is not found, the default output is given
Step 5: Providing a solution

B. CASE-BASED REASONING

Case-Based Reasoning (CBR) is a technique used to model human reasoning as well as an approach to building an intelligent system. CBR solves new problems by adapting solutions from similar cases. The process involved in CBR is represented by a four-step cycle in Fig. 1 [23]:

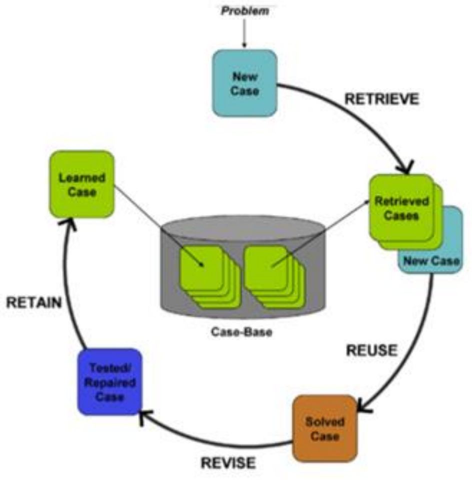

Figure 1. CBR step cycle

RETRIEVE – Retrieving a case from the databases that are similar to the new case. The new case matches the case found in the case database. This step is arguably a crucial phase where CBR performance is measured.

REUSE – Reusing the cases' solutions by copying or integrating the information and knowledge in that case to solve the new problem.

REVISE – Revising or adapting the proposed solution(s) to solve the new problem efficiently.

RETAIN – Retaining the new solution once it has been confirmed or validated that it is likely to be useful for future problem-solving.

C. THE PROPOSED CASE BASED REASONING APPROACH

This paper describes a comparative analysis of rule-based and case-based reasoning techniques using several commonly used similarity measures and a study on its performance for classification tasks. Also, we propose a new case-based reasoning approach using alternative similarity measure, called Distance-Weighted Case Base Reasoning (DW-CBR). The proposed method aims to improve classification performance. The following subsections describe the selected similarity measures.

1. Nearest Neighbor similarity measures

The nearest neighbor method is applied to calculate the similarity function and the total similarity (TS_i) of a potentially useful case [24].

$$sim\ (f_i^P, f_i^R) = 1 - |f_i^P - f_i^R| / \max(f_i) \quad (1)$$

$$TS_i = \left(\sum_{i=1}^{n} w_i * sim\ (f_i^P, f_i^R) \right) \Big/ \sum_{i=1}^{n} w_i \quad , (2)$$

where w_i is the weight of the ith attribute and $sim\ (f_i^P, f_i^R)$ is the function of the similarity of the ith attribute between the value of the new case f_i^P and the value of the retrieved case f_i^R.

2. Measures derived from Minkowski's Metric

$$d(C_{ik}, C_{jk}) = \left(\sum_{k=1}^{n} |C_{ik} - C_{jk}|^r \right)^{1/r} \quad r \geq 1 \quad , (3)$$

where n is the number of input attributes. When r =1, Manhattan Distance is obtained, and if r = 2, Euclidean distance is obtained. When including weights for all the attributes, the general formula becomes the following [15]:

$$d(C_i, C_j) = \left(\frac{\sum_{k=1}^{n} w_k^r * |d(C_{ik}, C_{jk})|^r}{\sum_{k=1}^{n} w_k^r} \right)^{1/r} \quad (4)$$

3. The proposed DW-CBR approach

The proposed DW-CBR approach uses the basic idea of close neighbor weighting according to their distance from the query [19]. We recommend DW-CBR using the dual distance-weighted function. In this new rule, we employ the dual distance-weights of nearest neighbors to determine the class of the query by majority weighted voting.

Denote the set $T' = \{(x_i^{NN}, y_i^{NN})\}_{i=1}^{k}$, arranged in increasing order in terms of Euclidean distance $d(x', x_k^{NN})$ between x' and x_k^{NN}.

$$Dw_i' = \begin{cases} \frac{d(x',x_k^{NN}) - d(x',x_i^{NN})}{d(x',x_k^{NN}) - d(x',x_1^{NN})}, if\ d(x',x_k^{NN}) \neq d(x',x_1^{NN}) \\ 1 \quad , if\ d(x',x_k^{NN}) = d(x',x_1^{NN}) \end{cases} \quad (5)$$

$$DWS_i = \left(\sum_{i=1}^{n} w_i * Dw_i'\right) \Big/ \sum_{i=1}^{n} w_i \quad (6)$$

Then, the result of the query is made by the majority weighted voting:

$$y' = \underset{y}{\arg max} \sum_{(x_i^{NN}, y_i^{NN}) \in T'} DWS_i\ x\ \delta(y = y_i^{NN}) \quad (7)$$

$\delta(y = y_i^{NN})$, the Dirac delta function takes a value of one if $y = y_i^{NN}$ and zero otherwise. According to the Eq. (5), a neighbor with a smaller distance is weighted more heavily than one with greater distance: the nearest neighbor gets a weight of 1, the furthest neighbor a weight of 0, and the other neighbors' weights are scaled linearly to the interval in between.

III. EXPERIMENTAL RESULT

The classification of early childhood diseases consists of four steps. The first step was pre-processing. The second step was normalization. The third step was the learning process, where the classification model was created. The third step was also called the learning or training phase. Data generated from this step was called the "training data". The four stages were classification, where the model was used for the first time to predict the data class labels. Classifier accuracy is seen from the percentage of test data that is correctly identified.

Data collection was done by reading patients' medical records – the data obtained from a hospital and a health center in Surabaya, Indonesia. The number of collected data was 1300, divided into training and testing data, 1040 and 260, respectively. The subset of data used in the experiment is publicly available in [25].

The first step was to pre-process the data. The data used was obtained from a hospital and a health center in Surabaya, Indonesia, and collected from interviews and documentation. The data of each patient consists of dates of treatment, age, sex, weight, height, body temperature, and any records of experiences of the 26 symptoms of 18 diagnoses. The patient's data are grouped based on [26], as seen in Table 1.

Table 1. The Patient's Grouping

Age (Month)	Age Feature	Weight (Kg)	Weight Feature	Height (cm)	Height Feature
0-12	U_1	3 – 9	B_1	49 – 76	T_1
13-24	U_2	10 – 12	B_2	77 – 87	T_2
25-36	U_3	13 – 14	B_3	88 – 96	T_3
37-48	U_4	15 – 16	B_4	97 – 103	T_4
49-60	U_5	17 – 19	B_5	104 – 110	T_5
61-72	U_6	19 – 24	B_6	111 – 116	T_6

The second step was to do normalization. Min-max normalization is then applied to transform a value v of a numeric attribute A to v' in the range of 0. This is defined by [27]:

$$v' = \frac{v - min_A}{max_A - min_A} \quad (8)$$

The third step was the learning process, where the classification model was created. This paper focus on the main technical characteristics of two reasoning techniques, which are commonly used, namely; rule-based reasoning and case-based reasoning.

A. RULE BASED REASONING

The third step was to create a classification model. The steps in applying rule-based reasoning were determined to use forward chaining. The data labels used in this study were 18 diagnoses and the 26 symptoms. There are three common symptoms, namely, cough, diarrhea, and fever [28]. Codification of common symptoms used in the rest of the paper was K1 = cough, K2 = diarrhea, and K3= fever. Data and information about the symptoms, diseases, and complaints experienced by the childhoods are shown in Table 2 and Table 3. Based on the collected data about the diseases, eighteen rules were generated, as can be seen in Table 4.

Table 2. Symptoms Variable

Code	Symptoms
G1	Inability to drink or suckle
G2	Vomiting
G3	Seizures
G4	Unconsciousness
G5	Dizziness
G6	Fast breathing
G7	Stridor
G8	liquid or soft defecating
G9	Hollowed eyes
G10	Poor abdominal skin turgor
G11	Fussiness/ irritability
G12	Abnormal thirst
G13	Restlessness
G14	Diarrhea for 14 days or more
G15	Blood in feces
G16	Paling
G17	Stiff neck (child cannot nod until chin touches chest)
G18	Red spots
G19	Red eyes
G20	Turbidity on the cornea
G21	Mouth ulcer
G22	Purulent eyes
G23	Fever for 2 to 7 days
G24	High and continuous sudden fever
G25	Heartburn
G26	Nausea

Table 3. Diseases Table

Code	Classification of disease
P1	Common risk sign
P2	Cough
P3	Pneumonia
P4	Severe Pneumonia
P5	Diarrhea
P6	Mild diarrhea with dehydration
P7	Diarrhea with severe dehydration
P8	Persistent diarrhea
P9	Severe persistent diarrhea
P10	Dysentery
P11	Common fever
P12	Severe fever
P13	Measles
P14	Measles with complication
P15	Measles with severe complication
P16	Fever that may be DHF
P17	DHF
P18	Fever that is not DHF

B. THE PROPOSED APPROACH

The next step was to create a classification model using the proposed case-based reasoning approach. Case representation was used to identify variables, namely weight, height, gender, age, body temperature, and the 26 symptoms. The complete list of symptoms can be seen in Table 2. Variables are grouped into two kinds, namely generic and specific variables. The former have general properties, which are height, weight, temperature, age, and sex, whereas the latter are variables distinct to particular diseases.

Table 4. Generated Rules

Rule	IF	THEN
1	G1 OR G2 OR G3 OR G4	P1
2	K1 AND G5	P2
3	K1 AND G6	P3
4	K1 AND P1 OR G7	P4
5	K2 AND G8	P5
6	P5 AND G10 AND G11 OR G12 OR G13	P6
7	P5 AND G10 OR G12 OR G13	P7
8	P5 AND G14	P8
9	P8 AND P6 OR P7	P9
10	P5 AND G15	P10
11	K3 AND G16	P11
12	P1 AND P11 OR G17	P12
13	P11 AND G18 AND G19 OR G21	P13
14	P13 AND P1 AND G21 OR G20	P14
15	P13 AND G21 OR G22	P15
16	P11 AND G23 AND G24	P16
17	P11 AND G23 AND G24 OR G25 OR G11	P17
18	P11 AND G16 OR G26	P18

Case-based reasoning begins by determining the case base, followed by expert weighting, finding local similarity, determining confidence level, finding global similarity, and selecting the highest value.

The first step in case-based reasoning is expert weighting. Weighting, which is the process of determining weights for each variable, is carried out by an expert. Weights are calculated using statistical analysis, starting with counting the number of variables that appear in the target case in the training data. Weighting is done through a series of iterations. There are two prerequisites for the calculation process. The first prerequisite is that the highest global similarity value results from the comparison value between the test data and the training data must match the doctor’s diagnosis data. If the highest global similarity value does not match the doctor’s diagnosis, the weight value of the diagnosis system must be adjusted to get a lower similarity value than the global similarity value obtained from the doctor’s diagnosis. Weights are adjusted using similarity of variables from testing and training data. If the variable similarity value is 1 and the weight is high, then the weight must be reduced. If the similarity value of the variable is 0, then the weight must be increased. The second prerequisite is the normalization of weights, which is completed by maintaining the total weight value of all diseases so that they have the same value. In this case, the total weight value of all diseases is set to 290.

The next step is to find a local similarity. Calculation of local similarity is the process of comparing the value of symptom variables between new cases and old cases. The local similarity calculation for CBR using NN used Eq. 1, and its derivative from Minkowski’s metric used Eq. 3 and the proposed DW-CBR used Eq.5.

After calculating local similarity, the next step is to calculate confidence value by comparing the number of variables in the new case and the previous case. The results of calculating the confidence value are used to decide whether the new case can proceed to the next calculation step, or should be pushed back to the previous step. The confidence threshold of all data is < 0.75. Anything lower than this has to repeat the process.

The next step is to find a global similarity. The algorithm used in the process of global similarity is the Nearest Neighbor similarity measure using Eq. 2, The Measure is derived from Minkowski’s Metric using Eq. 4, and the proposed DW-CBR using Eq. 6. The highest value is obtained from the calculation of global similarity, showing the diagnosis of new cases.

The revision process is carried out only if the confidence level calculation process has a value of < 0.75. This process is carried out in consultation with experts about cases that cannot be diagnosed by the system. Determining whether the system can or cannot answer a diagnosis comes from the level of confidence. The value is set at <0.75. This level of confidence means that the value does not resemble all data in the database.

If new cases are calculated with global similarity and receive a diagnosis, new case data will be entered into the database. The data is then considered as old data that will be compared with new data case sets.

C. EVALUATION AND DISCUSSION

The final step was to classify where the model was used for the first time to predict the data class labels. Classifier accuracy is seen from the percentage of test data that is correctly identified. To conduct a rigorous performance evaluation, sensitivity, and accuracy are taken into consideration [29]. Analysis of the results is carried out to determine whether the system that has been made is applicable in the diagnoses of early childhood diseases. The evaluation formula is as follows:

$$\text{Sensitivity} = \frac{T_P}{T_P + F_N} \text{ x } 100\% \qquad (8)$$

$$Accuracy = \frac{T_P}{T_P+F_P+T_N+F_N} \; x \; 100\ \%, \qquad (9)$$

where true positive (TP) is a case where we predict yes (they have the disease), and they do have the disease. True negative (TN) is we predict no, and they do not have the disease, false positive (FP) is we predict yes, but they do not really have the disease, and false negative (FN) is when we predict no, but they suffer from that disease.

This paper describes a comparative analysis of rule-based and case-based reasoning techniques using four different similarity measures and studies their performance for classification tasks. Rule-based reasoning used forward chaining and case-based reasoning used four similarity measures, i.e., Nearest Neighbor (NN-CBR), Euclidean Distance (ED-CBR), Minkowski Distance (MD-CBR), and Distance-Weighted (DW-CBR).

The first step in analyzing the system is to create a confusion matrix based on each similarity value. Then, sensitivity and accuracy are calculated using Eq. 8 and 9. The result of the confusion matrix can be seen in Table 5, and the evaluation results can be seen in Table 6. The result of classification accuracy can be seen in Fig. 2.

Fig. 2. shows that CBR accuracy is higher than rule-based reasoning. Forward chaining does not use distance calculations but adapts the approach. This causes a classification process that is not rigorous and, in turn, decreases classification accuracy. CBR uses old knowledge and is able to adapt new knowledge, hence the ability to support justification by considering examples from past cases [30]. CBR is like an expert system that always generates new rules to solve the problem. If there are similarities, they will be used as an experience to solve a new case with a little adaptation that fits the new case condition [31]. Also, the resulting rules can be corrected or modified to obtain better results to increase accuracy [32]. CBR paradigm is the best reasoning technique regarding the issues of maintenance and knowledge representations.

Table 5. Result of Confusion Matrix

Diseases	Early Childhood Diseases (Y) Prediction		Not Early Childhood Diseases (N) Prediction	
	Y	**N**	**Y**	**N**
Forward Chaining	162 (TP)	98 (FP)	0 (FN)	0 (TN)
NN-CBR	180 (TP)	80 (FP)	0 (FN)	0 (TN)
ED-CBR	210 (TP)	50 (FP)	0 (FN)	0 (TN)
MD-CBR	218 (TP)	42 (FP)	0 (FN)	0 (TN)
The proposed DW-CBR	240 (TP)	20 (FP)	0 (FN)	0 (TN)

Table 6. Evaluation Result

Type of Evaluation	Sensitivity (%)	Accuracy (%)
Forward Chaining	100	62
NN-CBR	100	69
ED-CBR	100	81
MD-CBR	100	82
The proposed DW-CBR	100	92

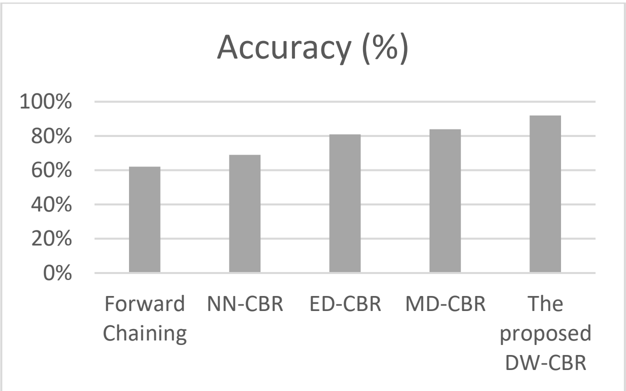

Figure 2. Result of Classification Accuracy

The proposed DW-CBR model produces better accuracy performance than other algorithms. The best accuracy of 92% is obtained by the proposed DW-CBR. The order of accuracy performance is as follows: The proposed DW-CBR > MD-CBR> ED-CBR> NN-CBR> Forward Chaining.

The proposed DW-CBR has the robustness to the sensitivity of different choices of the neighborhood. The proposed DW-CBR uses the basic idea of close neighbor weighting according to their distance from the query [19]. The proposed DW-CBR uses the dual distance-weighted function. In this new rule, a dual distance weighting of nearest neighbors is employed to determine the query class based on majority weighted voting.

IV. CONCLUSIONS

Research on early childhood diseases is essential because of the increasing mortality of early childhood each year. Reasoning techniques to diagnose and treat early childhood illnesses are needed to help specialist doctors and patients.

The area of research in this field covers rule-based and case-based reasoning. Case-based reasoning accuracy is higher than rule-based reasoning. The case-based reasoning paradigm is the best reasoning technique methodology regarding the problem of maintenance and knowledge representations.

Case-based reasoning uses four similarity measures, i.e., Nearest Neighbor (NN-CBR), Euclidean Distance (ED-CBR), Minkowski Distance (MD-CBR), and Distance-Weighted (DW-CBR). The best accuracy of 92% is obtained by the proposed DW-CBR. The proposed DW-CBR model has better accuracy performance than other algorithms. This is because the proposed DW-CBR has the robustness to the sensitivity of different choices of the neighborhood. The proposed DW-CBR uses the dual distance-weighted function. In this new rule, a double distance weighting of nearest neighbors is employed to determine the query class based on majority weighted voting.

In our future endeavors, we will attempt to investigate other alternative prediction models related to hybrid case-based reasoning. Hybrid case-based reasoning consists of combining artificial neural networks (ANN's) and CBR. ANN's are well-known, massively parallel computing models that exhibit desirable behavior in input-output mapping and resolving complex artificial intelligence problems in prediction and classification tasks.

References

[1] B. F. Yanto, I. Werdiningsih, and E. Purwanti, "Expert system application of early childhood diseases diagnosis using forward chaining method," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 3, no. 1, Indonesian version, pp. 61–67, 2017, https://doi.org/10.20473/jisebi.3.1.61-67.

[2] M. Garenne, C. Ronsmans, and H. Campbell, "The magnitude of mortality from acute respiratory infections in children under 5 years in developing countries.," *World Health Stat. Q.*, vol. 45, no. 2–3, pp. 180–191, 1992.

[3] Ministry of National Development, *Report on The Achievement of The Millennium Development Goals Indonesia 2011*. National Development Plannying Agency (BAPPENAS), 2012, pp. 1-16. (in Indonesian)

[4] N. H. Perlroth and C. W. Castelo Branco, "Current knowledge of environmental exposure in children during the sensitive developmental periods," *J. Pediatr. (Rio. J).*, vol. 93, no. 1, pp. 17–27, 2017, https://doi.org/10.1016/j.jped.2016.07.002.

[5] M. Fisun, M. Dvoretskyi, H. Horban, and M. Komar, "Knowledge management applications based on user activities feedback," *International Journal of Computing*, vol. 18, no. 1, pp. 32–44, 2019, https://doi.org/10.47839/ijc.18.1.1271.

[6] I. D. Wilson, S. D. Paris, J. A. Ware, and D. H. Jenkins, "Residential property price time series forecasting with neural networks," *Knowledge-Based Syst.*, vol. 15, no. 5–6, pp. 335–341, 2002, https://doi.org/10.1016/S0950-7051(01)00169-1.

[7] N. B. Chaphalkar and S. Sandbhor, "Use of artificial intelligence in real property valuation," *Int. J. Eng. Technol.*, vol. 5, no. 3, pp. 2334–2337, 2013.

[8] J. R. F. Clifton, "Applications of Computers and Information Technology," in *Handbook of Analytical Techniques in Concrete Science Technology*, 2001, pp. 1–2, https://doi.org/10.1016/B978-081551437-4.50021-7.

[9] A. Aamodt, "Case-based reasoning: Foundational issues, methodological variations, and system approaches, as presented by Praveen Guddeti," *AI Commun.*, vol. 7, no. 1, pp. 39–59, 1994, https://doi.org/10.3233/AIC-1994-7104.

[10] J. Kolodner, "An Introduction to Case-Based Reasoning," *Artif. Intell. Rev.*, vol. 6, no. 1, pp. 3–34, 1992, https://doi.org/10.1007/BF00155578.

[11] P. C. Lee, T. P. Lo, M. Y. Tian, and D. Long, "An efficient design support system based on automatic rule checking and case-based reasoning," *KSCE J. Civ. Eng.*, vol. 23, no. 5, pp. 1952–1962, 2019, https://doi.org/10.1007/s12205-019-1750-2.

[12] S.-T. Li and H.-F. Ho, "Predicting financial actifity with evolutionary fuzzy case based reasoning," *Expert*, vol. 36, no. 1, pp. 411–422, 2009, https://doi.org/10.1016/j.eswa.2007.09.049.

[13] S.-W. Lin and S.-C. Chen, "Parameter tuning, feature selection and weight assignment of features for case-based reasoning by artificial immune system," *Applied Soft Comput. J.*, vol. 11, no. 8, pp. 5042–5052, 2011.

[14] T. W. Liao, Z. Zhang, and C. R. Mount, "Similarity measures for retrieval in case-based reasoning systems.," *Appl. Artif. Intell.*, vol. 12, no. 4, pp. 267–288, 1998, https://doi.org/10.1080/088395198117730.

[15] H. Núñez *et al.*, "A comparative study on the use of similarity measures in case-based reasoning to improve the classification of environmental system situations," *Environ. Model. Softw.*, vol. 19, no. 9, pp. 809–819, 2004, https://doi.org/10.1016/j.envsoft.2003.03.003.

[16] H. Ahn and K. J. Kim, "Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach," *Appl. Soft Comput. J.*, vol. 9, no. 2, pp. 599–607, 2009, https://doi.org/10.1016/j.asoc.2008.08.002.

[17] P. D. Gutiérrez, M. Lastra, J. Bacardit, J. M. Benítez, and F. Herrera, "GPU-SME-kNN: Scalable and memory efficient kNN and lazy learning using GPUs," *Inf. Sci. (Ny).*, vol. 373, pp. 165–182, 2016, https://doi.org/10.1016/j.ins.2016.08.089.

[18] Y. Zeng, Y. Yang, and L. Zhao, "Pseudo nearest neighbor rule for pattern classification," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3587–3595, 2009, https://doi.org/10.1016/j.eswa.2008.02.003.

[19] J. Gou, L. Du, Y. Zhang, and T. Xiong, "A new distance-weighted k-nearest neighbor classifier," *J. Inf. Comput. Sci*, vol. 9, pp. 1429–1436, 2012.

[20] J. W. Moore and L. M. Quintero, "Comparing forward and backward chaining in teaching Olympic weightlifting," *J. Appl. Behav. Anal.*, vol. 52, no. 1, pp. 50–59, 2019, https://doi.org/10.1002/jaba.517.

[21] I. M. Ahmed, M. Alfonse, M. Aref, and A. B. M. Salem, "Reasoning Techniques for Diabetics Expert Systems," *Procedia Comput. Sci.*, vol. 65, pp. 813–820, 2015, https://doi.org/10.1016/j.procs.2015.09.030.

[22] E. Shortliffe and A. Scott, *An expert system for oncology protocol management*, William Kauman, Inc., I.os Ahos, 1984, pp 653-665.

[23] S. C. K. S. Sankar K. Pal, *Foundations of Soft Case-Based Reasoning*, A John Wiley & Son, Inc., New Jersey, 1993, pp. 11-18.

[24] M. Relich and P. Pawlewski, "A case-based reasoning approach to cost estimation of new product development," *Neurocomputing*, vol. 272, pp. 40–45, 2018, https://doi.org/10.1016/j.neucom.2017.05.092.

[25] I. Werdiningsih, "childhood diseases dataset," 2019. [Online]. Available: http://bit.ly/childhooddiseases.

[26] Indonesian Government of Health, "*Decision from Health Ministery of Indonesian Government Number 1995 Year 2010 About Antropometry Standart on Evaluating Children Nutrition*". Indonesian Government of Health, Jakarta, Indonesian version, 2101, pp. 5–40.

[27] Z. E. Rasjid and R. Setiawan, "Performance comparison and optimization of text document classification using k-NN and Naïve Bayes classification techniques," *Procedia Comput. Sci.*, vol. 116, no. December, pp. 107–112, 2017, https://doi.org/10.1016/j.procs.2017.10.017.

[28] Indonesian Government of Health, *Manajemen Terpadu Balita Sakit (MTBS)*, Indonesian Government of Health, Jakarta, Indonesian version, 2011, pp 1-8.

[29] I. H. Witten, E. Frank, and M. a Hall, *Data Mining:Practical Machine Learning Tools and Techniques second edition*, Morgan Kaufmann, Inc., San Fransisco, 2011, pp. 70-76.

[30] I. Watson, *Applying Case-Based Reasoning: Techniques for Enterprise Systems (The Morgan Kaufmann Series in Artificial Intelligence)*, Morgan Kaufmann, Inc., San Fransisco 1997, pp.1-6.

[31] R. Qu, *Case-Based Reasoning for Course Timetabling Problems*, University of Nottingham, 2002, pp. 10-38.

[32] S. T. Deepa and S. G. Packiavathy, "Expert system for car troubleshooting," *Gener. 5*, vol. 1, no. 1, pp. 46–49, 2012.

***INDAH WERDININGSIH** graduated from Sepuluh Nopember Institute of Technology in 2011 with speciality "Computation and Computer Vision". She worked in Surabaya University as a Lecturer in Computer Science Department from 2003 to 2004. At the moment she is Senior Lecturer of Mathematics Department, Information Systems Study Program, Faculty of Science and Technology, University of Airlangga. Areas of scientific interests include: Artificial Intelligent, Expert System, and Data mining. Indah is a member of the Association for Information Systems – Indonesia Chapter, and Indonesian Computer Science Higher Education Association (APTIKOM).*

***RIMULJO HENDRADI** received his B.Sc. and M.Sc. degrees in Mathematics, University of Gadjah Mada, in 1995 and 2002, respectively. He received Ph.D. degree at Electrical Engineering, Institute of Technology Sepuluh Nopember in 2016. He is presently an Assistant Professor at Mathematics Department, Information Systems Study Program, Faculty of Science and Technology, University of Airlangga.*

His main research interests include the biomedical signal processing and particular in time-frequency phonocardiogram analysis, also decision support systems. Dr. Hendradi is a member of the IEEE Engineering in Medicine and Biology Society, Association for Information Systems – Indonesia Chapter, and Indonesian Computer Science Higher Education Association (APTIKOM).

***PURBANDINI** graduated from Sepuluh Nopember Institute of Technology in 2007 with speciality "Data Mining". She worked in a company related with data analysis in 1993 to 2003. Then, since 2007 until 2008 she acted as Head of Computer Science Department in Muhammadyah University, Jember. At the moment she is a Senior Lecturer of the Department of Information System of the Universitas Airlangga. Areas of scientific interests include: Artificial Intelligent, Data Mining, Geographic Information Systems, Decision Support System. Purbandini is a member of the Association for Information Systems – Indonesia Chapter, and Indonesian Computer Science Higher Education Association (APTIKOM).*

***BARRY NUQOBA** graduated from Sepuluh Nopember Institute of Technology in 2009 with speciality "Data Mining". At the moment he is a Lecturer of the Department of Information System of the Universitas Airlangga. Areas of scientific interests includes: Artificial Intelligent, Data Mining, Cyber-Physical System. Barry is a member of the Association for Information Systems – Indonesia Chapter, and Indonesian Computer Science Higher Education Association (APTIKOM).*

***ELLY ANNA**, graduated from Bogor Agricultural Institute 2000 with speciality "Statistics". At the moment she is an Associated Professor of Mathematics Department, Statistics study Program, University of Airlangga. Her main research interests include Statistics techniques, Multivariate methods, and Data mining.*