Determination of children's nutritional status with machine learning classification analysis approach

Musli Yanto[1], Febri Hadi[1], Syafri Arlis[2]
[1]Department of Informatics Engineering, Faculty of Computer Science, Universitas Putra Indonesia YPTK, Padang, Indonesia
[2]Department of Information System, Faculty of Computer Science, Universitas Putra Indonesia YPTK, Padang, Indonesia

Article Info

Article history:

Received Jul 18, 2023
Revised Nov 1, 2023
Accepted Nov 4, 2023

Keywords:

Analysis model
Classification
Knowledge based system
Machine learning
Malnutrition

ABSTRACT

Malnutrition is a problem that is often faced by every country around the world. Various facts show that malnutrition is of particular concern to many researchers. To can overcome this problem, every effort has been made such as developing analytical models in identification, classification, and prediction. This study aims to determine the nutritional status of children using the machine learning (ML) classification analysis approach. The methods used in the ML analysis process consist of cluster K-Means, artificial neural network (ANN), sum square error (SSE), pearson correlation (PC), and decision tree (DT). The dataset for this study uses data on child nutrition cases that occurred in the previous and was sourced from the provincial general hospital (RSUP) M. Djamil, Padang, West Sumatera. Based on the research presented, ML performance in the nutritional status classification analysis gave maximum results. These results are reported based on the level of precision with an accuracy of 99.23%. The results of the analysis can also present a knowledge-based nutritional status classification. This research can contribute to and update the analytical model in determining nutritional status. The results of this study can also provide benefits in handling nutritional status problems that occur in children.

Corresponding Author:

Musli Yanto
Department of Informatics Engineering, Faculty of Computer Science, Universitas Putra Indonesia YPTK
25145 Lubuk Begalung, Padang, Indonesia
Email: musli_yanto@upiyptk.ac.id

1. INTRODUCTION

Malnutrition is a case of imbalance that occurs between food intake and energy needs in the human body [1]. Cases of malnutrition can affect anyone, including children, adolescents, adults and even the elderly [2]. Malnutrition cases also provide gaps or opportunities to be attacked by various diseases such as stunting, anemia, and others [3], [4]. This case is also indicated as a result of a lack of information in understanding the causes of malnutrition [5]. So with this it is necessary to have a process of analyzing malnutrition status which is measured based on indicators determining the level of nutritional status [6].

Indicators in determining nutritional status basically use parameters of human size, shape, and proportion [7]. Previous research also explained that nutritional status is determined by taking physical measurements based on body mass index (BMI) [8]. Furthermore, physical examination or what is known as the anthropometric approach including measuring height and weight is an indicator in determining malnutrition status [9]. The form of measurement by calculating the weight index based on age, weight based on height and height based on age through the Z-Score is a process in determining nutritional status [10].

With this explanation, the determination of nutritional status can be implemented into the classification process. Previous research explained that the classification process for malnutrition status using the K-Nearest Neighbor algorithm gave a classification result of 85.24% [11]. The application of the data mining classification analysis model is also able to predict nutritional status by producing solutions that can be used as information [12]. The application of other data mining concepts has also been developed in determining nutritional status using datasets from the results of physical measurements [13]. Data mining techniques in classification and clustering analysis can predict the nutrition of children under 5 years based on height and weight according to age [14]. The development of a classification analysis model in other forms has also been developed with the concept of an expert system using the forward chaining method in determining nutritional status [15]. So that in this case the classification analysis process based on the indicators used gives effective results in determining nutritional status [16].

Based on the history of previous research, this study will also carry out a classification analysis process in determining nutritional status. The analysis process is carried out by developing an analytical model of the machine learning (ML) approach. Previous studies explained that the application of ML is used to predict malnutrition status in toddlers [17]. ML is used as a tool to assist humans in making decisions [18]. ML performance results in a good level of accuracy in the identification process using supervised concepts [19]. Other experimental results show that ML by classifying an intelligent system has good results compared to other methods [20]. The ML classification process presents a model that is able to evaluate the learning process against a data set [21]. Based on previous ML research reports, this study proposes to optimize ML performance in the process of analyzing nutritional status classification. The goal achieved is to develop a classification analysis model to present much better results than before. The development of the ML classification analysis model is presented at the initial pre-processing stage and the learning process.

The pre-processing development stage will produce an optimal analysis pattern. The pattern is generated by conducting tests to measure the strength of each indicator variable using the sum square error (SSE) and pearson correlation (PC) methods. SSE performance can calculate the error rate based on the data pattern formed [22]. Not only that, but SSE is also able to provide stability in analyzing patterns in classifying [23]. In addition to SSE, the PC method also plays an important role in measuring the performance of analytical indicators [24]. The PC method has been proven to be able to present a correlation level relationship between variables [25]. With the performance of the SSE and PC methods, the resulting classification analysis pattern will be able to provide an optimal pattern in determining nutritional status.

The ML learning process stage was developed to present precise and accurate analysis results. Learning outcomes can support the performance of the analysis model in classifying. The ML learning process uses the artificial neural network (ANN) method and the decision tree (DT) method. In principle, ANN is one method that is widely used to carry out the classification process. ANN is also used in solving complex problems by imitating how the human brain works [26]. The application of ANN in carrying out classification analysis has been able to provide a fairly good contribution [27]. ANN adopts a learning process using a modified feedforward algorithm based on network architectural patterns [28]. Not only the ANN method and the DT method are also proposed to provide an overview of decision trees in determining nutritional status. The DT method is capable of presenting knowledge-based patterns of analysis [29]. DT performance can turn facts into information and knowledge [30]. Overall the performance of the DT method in the classification analysis process will provide maximum results [31].

Based on this explanation, this research aims to develop an ML classification analysis model for determining nutritional status. The development of this analytical model is aimed at presenting an effective and efficient classification process. The analytical model developed will also provide updates to the analysis process on ML performance. This update can have an impact on the performance of ML learning by modifying an algorithm. The results of ML development can later be expected to provide precise and accurate output based on a fairly good level of accuracy. Overall, this research can provide contributions and benefits in handling and control in monitoring the development of nutritional status cases for the hospital and other interested parties.

2. METHOD

In carrying out the classification analysis process for determining nutritional status in children, the concept of ML is adopted to provide maximum output. Machine learning has a pretty good performance in dealing with classification analysis problems [32]. Machine learning also provides a fairly good description of the analytical model in the case of classification [33]. Machine learning can be applied in a model of this research framework in carrying out the analysis process of determining nutritional status. The description of the research framework can be presented in Figure 1.

Figure 1 is a research framework presented in the model for carrying out classification analysis. This model was developed to maximize ML performance to provide optimal analysis results. ML development performance is presented in 2 process stages, preprocessing and classification analysis. In more detail, the process stages can be explained as:

- Preprocessing analysis: The preprocessing analysis stage aims to produce the best analysis pattern in determining nutritional status. This process is presented in data clustering with the performance of the K-Means algorithm which is optimized using the sum square error (SSE) and pearson correlation (PC) methods. K-Means optimization was carried out to test the resulting analysis pattern based on the error rate and correlation between the analysis variables. The analysis pattern produced in preprocessing will maximize the performance of classification analysis in determining nutritional status.
- Classification analysis: The classification process in determining nutritional status will play the role of the ANN and DT methods in presenting the best output results. The analysis process will carry out learning based on previously obtained analysis patterns. The learning process is presented at the ANN training and testing stages as a form of classification analysis process. DT will also contribute actively in presenting analysis output which can be used as a knowledge-based system for determining nutritional status. Overall, the development of this analytical model can present a new algorithm for the classification process for determining nutritional status.

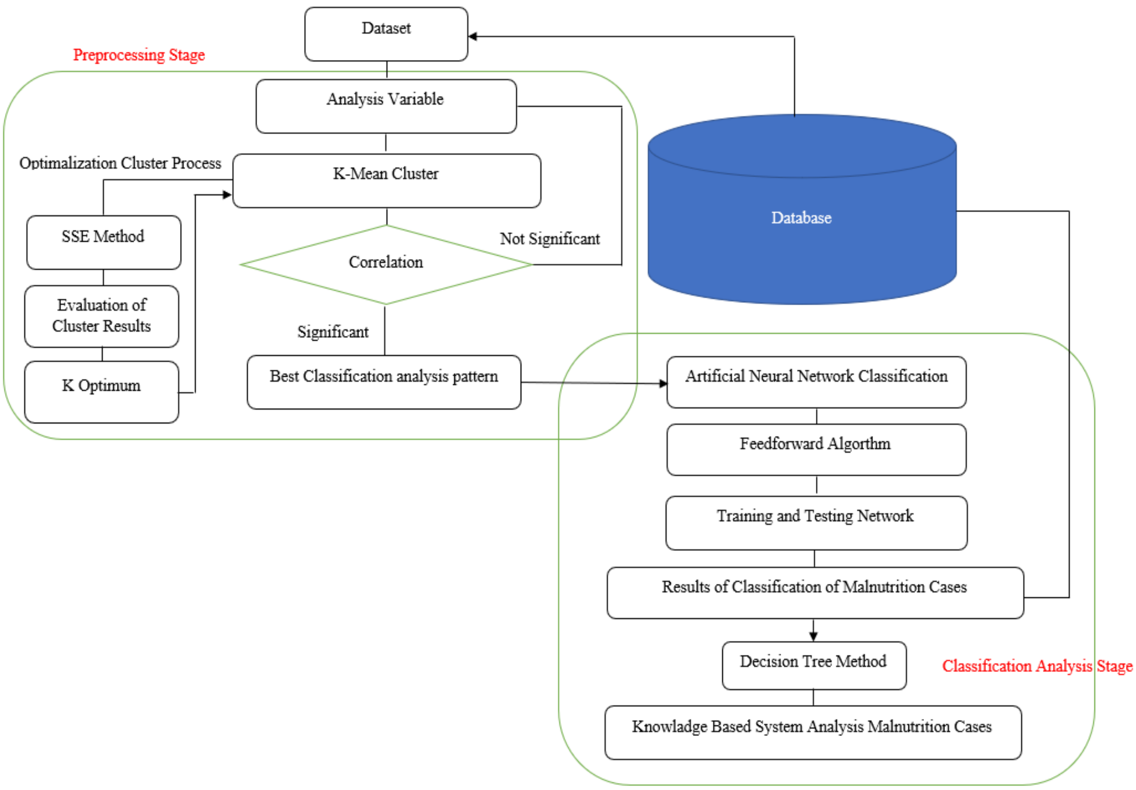


Figure 1. Research framework

2.1. Dataset

The research dataset in determining nutritional status uses data from medical examination results on children sourced from the M. Djamil Padang Provincial General Hospital (RSUP) in 2022. The analysis variables used in the dataset are Gender (X1), Age (X2), Weight (X3), and Height (X4). The sample research dataset can be presented in Table 1.

Table 1 is a presentation of the research dataset in the classification process to determine nutritional status. The dataset includes 20 samples from the 576 total data used in this research. The entire dataset will later be involved in the classification process in determining nutritional status.

Table 1. Sample research dataset

Patient	Gender	Age	Tall body	Weight
Patien 1	P	19	47.1	4.4
Patien 2	P	6	32	3.6
……..	..	..	..	..
Patien 20	P	4	36	2.8

2.2. K-Mean cluster

The basic concept of this method is very simple determining the number of clusters (k) and continuing to calculate the distance between each data [34]. The K-Mean performance process groups data into clusters based on the specified distance [20]. In the resulting performance, K-Means presents the cluster center value by calculating the new centroid until it is found that the value has not changed [35]. The working concept of the cluster method in calculating the distance from each data to the cluster center is presented in (1).

$$d(x,y) = \left||x-y|\right| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}\ ; i = 1,2,3, \ldots, n \tag{1}$$

According to (1) explains that d (x,y) is the process of calculating the centroid point with the point of each object using euclidean distance. Each data will later be measured using the euclidean distance in determining the cluster point. The results presented by the cluster method can form an analysis pattern.

2.3. Sum square error

The SSE method calculates the accuracy of distance measurements in calculating the total difference from the actual value. The SSE value can be used as a parameter in measuring the error level in a pattern [36]. Calculate the centroid value of the SSE, can be seen from (2) [37].

$$SSE = \sum_{i=0}^{k}(\mathrm{Xi} - \mathrm{Yi})^2 \tag{2}$$

According to (2) explains that SSE is a calculation to assess the error of an analysis process. The X value is the actual value and the Y value is the value to be achieved. SSE method used to measure the difference between the data obtained with the prediction model that has been done before. SSE is often used as a research reference in determining optimal clusters [38].

2.4. Person correlation

Pearson correlation is a statistical concept that is capable of performing calculations in carrying out a measurement process [39]. PC can be combined with several methods to provide better performance results [40]. PC-based techniques can also be used to select optimized features in reviewing output from a model [41]. The PC calculation can be presented in (3) [42].

$$P\ X,Y = \frac{\mathrm{E\ (XY)} - \mathrm{E\ (X)E\ (Y)}}{\sqrt{\mathrm{E\ (X.X)} - E.E(X)}\ \sqrt{B\ (Y.Y) - E.E\ (Y)}} \tag{3}$$

According to (3) explains that the value of cov (X, Y) is the covariance between X and Y. The value of X, Y is the standard deviation value of the variables X and Y. The value of E (X) is an expected value of X [42].

2.5. Artificial neural networks

ANN is a branch of science with the concept of an information processing paradigm inspired by the biological nervous system, such as information processing in the human brain. ANN is a method that can be used in carrying out the classification process by presenting a fairly good level of performance [43]. ANN performance provides the maximum level of results in dealing with problems such as identification, classification, and prediction [44], [45]. Based on the concept that has been explained, ANN can carry out learning by adopting a series of mathematical calculation processes [46]. Learning can be modified in a model that is applied in the form of an algorithm to produce decisions [47]. The model is presented in an architectural pattern based on the input layer, hidden layer, and output layer [48]. Overall the concept of ANN aims to provide optimal output from the learning process carried out [49].

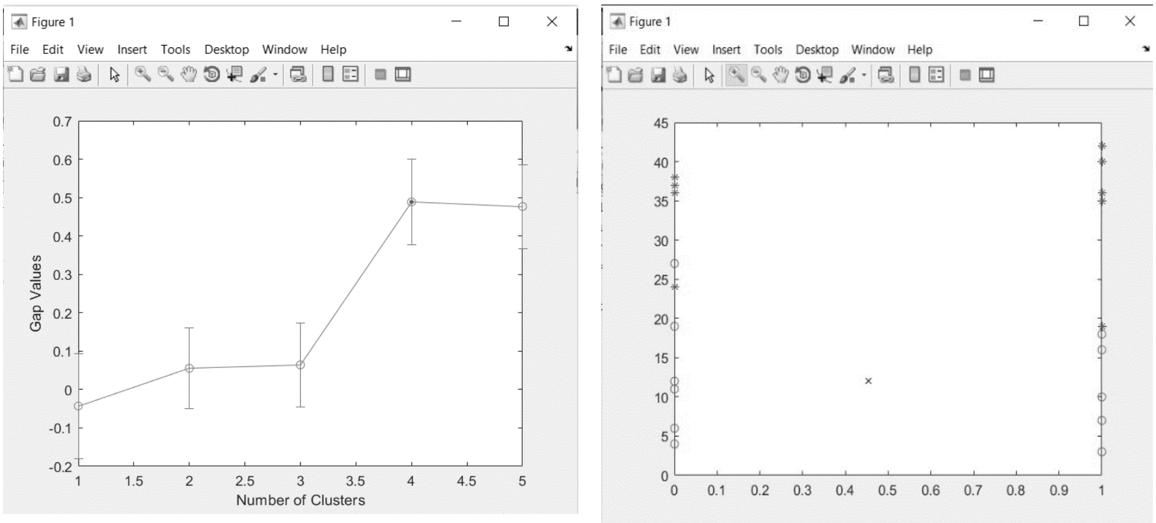
2.6. Decision tree

Previous research reviews describe the decision tree method used to produce a decision tree. The results of DT performance are used to train machine learning by producing decision trees that produce a classification of 93% accuracy values obtained [50]. In another discussion, the DT method is also applied to supervised machine learning [51]. The performance of the DT method provides output with an accuracy rate of 99% [52].

3. RESULTS AND DISCUSSION

The process of classification analysis in determining nutritional status is carried out by developing the preprocessing analysis stage and the learning analysis stage. This development can present maximum output results. In the preprocessing analysis stage, development is carried out by optimizing the performance of cluster processes in producing analysis patterns. The optimization process is presented in two directions, namely determining the optimal K value and measuring the error rate using the SSE method. The results of the preprocessing analysis experiments that have been carried out can be presented in Figure 2.

Figure 2 is the result of preprocessing analysis experiments to produce classification analysis patterns. Based on these tests, it can be seen that Figure 2(a) presents the results of experiments to determine the optimal K value. Optimal K value = 3 based on the resulting elbow graph. Figure 2(b) presents the cluster results based on the K values obtained. Apart from that, Figure 2(c) also presents the error rate of cluster processing results for each K value with SSE calculations. Based on these results, it can also be seen that the SSE value with K=3 has a minimal error rate among other experiments. Thus, the results of the clustering process can be presented in Table 2.



(a) (a)

SSE				
43207280	14296696	9281941	14662931	13720512

(c)

Figure 2. Preprocessing analysis results presented include: (a) K-mean elbow graph, (b) cluster analysis results, and (c) SSE calculation results

Table 2. Cluster analysis results

No	Gender	Age	Tall body	Weight	Status
1	F	19	47.1	4.4	Good nutrition
2	F	6	32	3.6	Good nutrition
3	M	7	30	4.2	Good nutrition
4	F	4	36	2.8	Good nutrition
5	F	11	51	5.6	Good nutrition
6	M	3	41	3.6	Good nutrition
7	M	10	35	3.3	Good nutrition
8	M	18	45	3.7	Good nutrition
9	M	40	55	5.1	Malnutirition
10	M	42	61	7.2	Malnutirition
11	M	19	68.7	8.2	Malnutirition
12	F	27	34.2	4.2	Good nutrition
13	F	36	57.9	5.1	Malnutirition
14	F	37	67.7	7.2	Malnutirition
15	M	16	41	3.7	Good nutrition
16	M	35	45	17	Malnutirition
17	M	36	60	4.2	Malnutirition
18	F	38	67.9	8.1	Malnutirition
19	M	50	66	59	Obesity
20	M	35	60	62	Obesity

Table 2 presents the results of the cluster analysis which will become a pattern of classification analysis in determining nutritional status. The analysis pattern will later be measured again to ensure the accuracy of the analysis variables. Measurement of pattern analysis involves the performance of the PC method as the final stage of the development of preprocessing classification analysis. The performance results of the PC method can be presented in Table 3.

Table 3. Pearson correlation analysis results

		Correlations				
		Gender	Age	Tall_body	Month	Case
Gender	Pearson correlation	1	0.625	0.654	0.792	0.587
	Sig. (2-tailed)		0.599	0.820	0.212	0.220
	N	20	20	20	20	20
Age	Pearson correlation	0.625	1	0.749**	0.487*	0.707
	Sig. (2-tailed)	0.599		0.000	0.029	0.978
	N	20	20	20	20	20
Tall_body	Pearson correlation	0.654	0.749**	1	0.393	0.773
	Sig. (2-tailed)	0.820	0.000		0.087	0.244
	N	20	20	20	20	20
Month	Pearson correlation	0.792	0.487*	0.393	1	0.715
	Sig. (2-tailed)	0.212	0.029	0.087		0.363
	N	20	20	20	20	20
Case	Pearson correlation	0.587	0.707	0.773	0.715	1
	Sig. (2-tailed)	0.220	0.978	0.244	0.363	
	N	20	20	20	20	20

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Table 3 is the result of measuring the PC method to ensure the correlation between the analysis variables in the classification of nutritional status. Based on these results it can be seen that the Gender variable (X1) has a correlation level of 0.587 to Case (Y). Age variable measurement (X2) also has a correlation level of 0.707 to Case (Y). furthermore the Tall_Body variable (X3) also has a fairly good correlation of 0.773 with Case (Y). The Month variable (X4) in the analysis results is also able to present a correlation of 0.715 to Case (Y). Thus it can be concluded that the analysis variables used have contributed to the determination of nutritional status. based on the measurement results of the PC method, the analysis pattern can be used as a pattern in the classification process. This pattern will later be forwarded to the process of learning analysis stages using the ANN method. In the ANN analysis process, the process begins with the architectural design of the analysis pattern. The ANN analysis pattern architecture can be seen in Figure 3.

Figure 3 explains that the architecture of the ANN analysis pattern has several interconnected layers including the input layer, hidden layer, and output layer. The input layer consists of 4 input units sourced from the previously used dataset indicators. The input data will later be carried out by a learning process by adopting a feedforward algorithm to get maximum results. The learning outcomes of ANN in the classification analysis process can be presented in Figure 4.

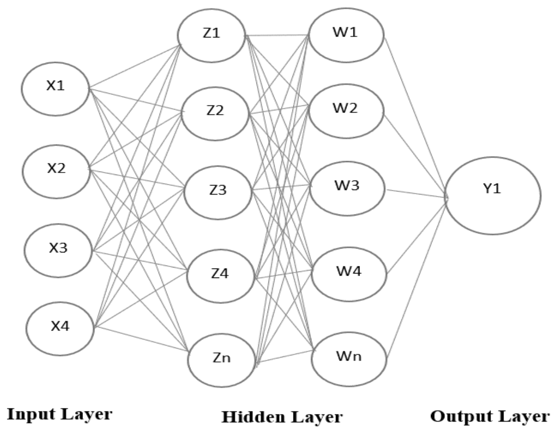


Figure 3. ANN analysis pattern architecture

Figure 4 is the output of classification analysis results based on the level of validation of the performance of the feedforward learning algorithm. The ANN output results will later be compared with several other output results based on network architecture pattern experiments. The comparison process is carried out to determine the best ANN architectural pattern. The results of this comparison can be presented in Table 4.

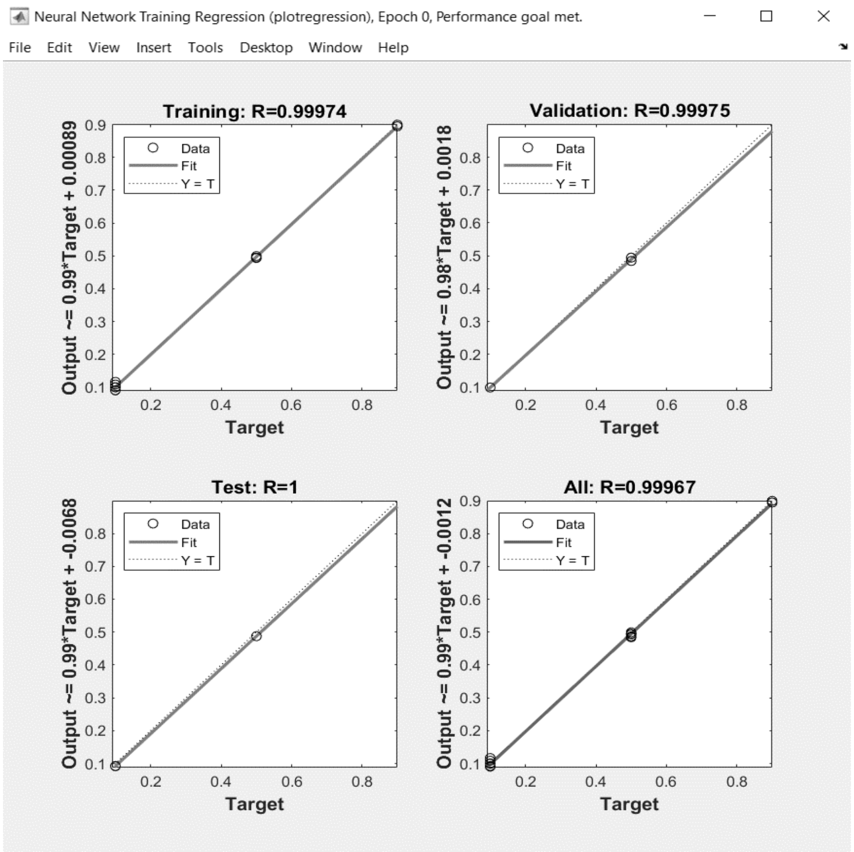


Figure 4. ANN learning validation results

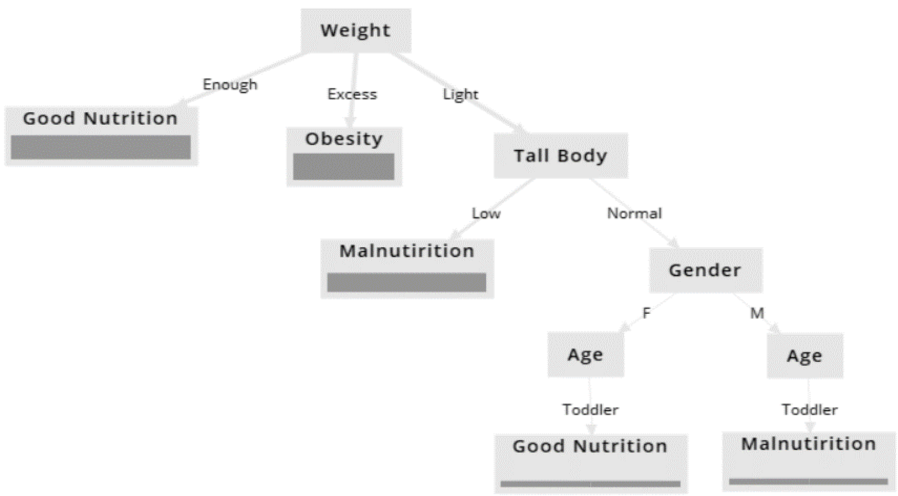
Table 4. Determination of the best architectural pattern

Architecture	Training			Testing		
Single hidden layer						
	Accuracy	MSE	Performance	Accuracy	MSE	Performance
(4-5-1)	99.730700	0.269300	0.002693	99.366950	0.633050	0.006331
(4-8-1)	99.998400	0.001600	0.001000	99.998400	0.001600	0.001000
(4-10-1)	99.717400	0.282600	0.002826	99.988600	0.011400	0.003300
(4-15-1)	99.693400	0.306600	0.001000	99.988500	0.011500	0.007200
(4-20-1)	98.420400	1.579600	0.001000	99.999446	0.000554	0.003700
Multi hidden layer						
Architecture	Training MSE			Testing		
	Accuracy	MSE	Performance	Accuracy	MSE	Performance
(4-5-5-1)	99.944100	0.055900	0.0016	99.997600	0.002400	0.001600
(4-5-8-1)	**99.444500**	**0.555500**	**0.000845**	**99.998600**	**0.001400**	**0.000845**
(4-5-5-8-1)	99.893900	0.106100	0.006300	99.997100	0.002900	0.003800
(4-5-8-8-1)	99.995800	0.004200	0.001400	99.998200	0.001800	0.001400
(4-5-8-10-1)	99.959600	0.040400	0.008600	99.959600	0.040400	0.004100

Table 4 presents the results of a comparison in determining the best architectural pattern in carrying out a classification analysis. The best architectural pattern has two hidden layers with each unit including 5 and 8. Thus the architectural pattern can be said to be multi-layer hidden. The results of the tests that have been carried out show that the output of the analysis gives an accuracy rate of 99.23% and a mean square error (MSE) value of 0.77%. Based on the test results, it can be concluded that the ANN classification analysis process has presented optimal output in determining nutritional status.

The classification process for determining nutritional status is continued with the analysis stage using the DT method. The performance of the DT method can present output results in the form of knowledge-based. These results will be reflected in a decision tree in dealing with determining nutritional status. The results of the analysis process with the DT method can be presented in Figure 5.

Figure 5 is the result of the analysis of the DT method in determining nutritional status. The results of the experiments that have been carried out get the output of a decision tree graph which is the knowledge base of classification analysis. The graph is also equipped with rules for determining nutritional status based on the resulting rule base. Based on the overall results of the classification analysis experiment with the ML approach, this study has been able to report optimal results of the analysis. These results can be measured with a maximum level of accuracy and minimal error.



Tree

```
Weight = Enough: Good Nutrition {Good Nutrition=9, Malnutirition=0, Obesity=0}
Weight = Excess: Obesity {Good Nutrition=0, Malnutirition=0, Obesity=10}
Weight = Light
|   Tall Body = Low: Malnutirition {Good Nutrition=0, Malnutirition=7, Obesity=0}
|   Tall Body = Normal
|   |   Gender = F
|   |   |   Age = Toddler: Good Nutrition {Good Nutrition=1, Malnutirition=1, Obesity=0}
|   |   Gender = M
|   |   |   Age = Toddler: Malnutirition {Good Nutrition=1, Malnutirition=1, Obesity=0}
```

Figure 5. Results of the DT method analysis process

Based on the discussion that has been carried out, the results of this research have been able to present an effective classification analysis process in determining nutritional status. The analytical performance developed has also been able to provide novelty in the ML classification analysis model. This novelty actively contributes to and has a significant impact on ML learning. The results of the analysis output will be used by related parties in handling and monitoring the growth of malnutrition that occurs in children.

4. CONCLUSION

The process of determining nutritional status using the ML approach is able to provide the maximum classification analysis process with the resulting output. This output has been proven to be precise and accurate in determining nutritional status, having an accuracy of 99.23%. based on these results it can be concluded that ML works with a pretty good performance. ML performance is based on the development of an analytical model that presents a novel classification process. This novelty can be seen from the initial preprocessing performance to produce analysis patterns. Another novelty is also presented in the classification process by presenting a knowledge-based system for determining nutritional status based on a decision tree. With the novelty generated in an analysis model, it can be used as an alternative solution in dealing with the problem of determining nutritional status. The resulting classification analysis model is also capable of playing an active role as a media control for handling nutrition cases.

ACKNOWLEDGEMENTS

Thanks to Mrs. Dr. Hj. Zerni Melmusi, SE, MM, Ak, CA as Yayasan Perguruan Tinggi Komputer (YPTK) and Prof. Dr. Sarjon Defit, S.Kom, M.Sc as Chancellor of Universitas Putra Indonesia YPTK Padang who has provided the opportunity for this research. This research was fully funded by Universitas Putra Indonesia YPTK Padang with agreement Nomor: 124/UPI-YPTK/LPPM/P/KP/VII/2023 funding for the 2023 Simlitupi Research.

REFERENCES

[1] S. Nakahara *et al.*, "Aggressive nutrition therapy in malnutrition and sarcopenia," *Nutrition*, vol. 84. 2021, doi: 10.1016/j.nut.2020.111109.

[2] N. K.-Braun, M. Mueller, P. Schuetz, B. Mueller, and A. Kutz, "Evaluation of nutritional support and in-hospital mortality in patients with malnutrition," *JAMA Network Open*, vol. 4, no. 1, 2021, doi: 10.1001/jamanetworkopen.2020.33433.

[3] S. Ghosh, "Factors responsible for childhood malnutrition: A review of the literature," *Current Research in Nutrition and Food Science*, vol. 8, no. 2. pp. 360–370, 2020, doi: 10.12944/CRNFSJ.8.2.01.

[4] M. O. Folayan *et al.*, "Associations between early childhood caries, malnutrition and anemia: A global perspective," *BMC Nutrition*, vol. 6, no. 1. 2020, doi: 10.1186/s40795-020-00340-z.

[5] A. S. Mustafa and W. R. Baiee, "Nutrition information estimation from food photos using machine learning based on multiple datasets," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 5, pp. 2922–2929, 2022, doi: 10.11591/eei.v11i5.4007.

[6] J. A. Nweze, E. I. Nweze, and U. S. Onoja, "Nutrition, malnutrition, and leishmaniasis," *Nutrition*, vol. 73. 2020, doi: 10.1016/j.nut.2019.110712.

[7] I. Sommer *et al.*, "The performance of anthropometric tools to determine obesity: a systematic review and meta-analysis," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020, doi: 10.1038/s41598-020-69498-7.

[8] P. Sharma *et al.*, "Nutritional assessment and factors affecting dietary intake in patients with cirrhosis: A single-center observational study," *Nutrition*, vol. 84, p. 111099, Apr. 2021, doi: 10.1016/j.nut.2020.111099.

[9] J. J. H.-Morante, C. Piernas, D. G.-Martínez, A. P.-Caballero, M. J. F.-Abellán, and I. M.-Moreno, "Health status and nutritional development of adopted Ethiopian children living in southern Spain: A prospective cohort study," *Nutrition*, vol. 71, p. 110611, 2020, doi: 10.1016/j.nut.2019.110611.

[10] A. Chissaque *et al.*, "Rotavirus A infection in children under five years old with a double health problem: undernutrition and diarrhoea – a cross-sectional study in four provinces of Mozambique," *BMC Infectious Diseases*, vol. 21, no. 1, pp. 1–13, 2021, doi: 10.1186/s12879-020-05718-9.

[11] S. Sendari, T. Widyaningtyas, and N. A. Maulidia, "Classification of toddler nutrition Status with anthropometry using the K-nearest neighbor method," in *ICEEIE 2019 - International Conference on Electrical, Electronics and Information Engineering: Emerging Innovative Technology for Sustainable Future*, 2019, vol. 6, pp. 154–158, doi: 10.1109/ICEEIE47180.2019.8981408.

[12] E. Oktavianti, A. R. Yuly, and F. Nugrahani, "Implementation of Naïve Bayes classification algorithm on infant and toddler nutritional Status," in *Proceedings - 2019 2nd International Conference of Computer and Informatics Engineering: Artificial Intelligence Roles in Industrial Revolution 4.0, IC2IE 2019*, 2019, pp. 170–174, doi: 10.1109/IC2IE47452.2019.8940894.

[13] Ermatita, S. Destriatania, and Yulnelly, "Nutrition anthtropometric status model by data mining: Case study in Palembang South Sumatera," *International Journal of Engineering Trends and Technology*, no. 1, pp. 97–103, 2020, doi: 10.14445/22315381/CATI2P215.

[14] Z. Markos, "Predicting under nutrition status of under-five children using data mining techniques: The case of 2011 Ethiopian demographic and health survey," *Journal of Health & Medical Informatics*, vol. 5, no. 2, 2014, doi: 10.4172/2157-7420.1000152.

[15] M. Ula, A. Fa. Ulva, I. Saputra, M. Mauliza, and I. Maulana, "Implementation of machine learning using the K-nearest neighbor classification model in diagnosing malnutrition in children," *Multica Science and Technology (Mst) Journal*, vol. 2, no. 1, pp. 94–99, 2022, doi: 10.47002/mst.v2i1.326.

[16] D. D. Salmarini and A. Citra S, "Nutrition detection status Of children based on the provision of breastmilk and supplement," 2017, doi: 10.2991/smichs-17.2017.7.

[17] A. Talukder and B. Ahammed, "Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh," *Nutrition*, vol. 78, p. 110861, 2020, doi: 10.1016/j.nut.2020.110861.

[18] L. N. Bonnell, B. Littenberg, S. R. Wshah, and G. L. Rose, "A machine learning approach to identification of unhealthy drinking," *Journal of the American Board of Family Medicine*, vol. 33, no. 3, pp. 397–406, 2020, doi: 10.3122/jabfm.2020.03.190421.

[19] S. M. Alotaibi, A. U. Rahman, M. I. Basheer, and M. A. Khan, "Ensemble machine learning based identification of pediatric epilepsy," *Computers, Materials and Continua*, vol. 68, no. 1. pp. 149–165, 2021, doi: 10.32604/cmc.2021.015976.

[20] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart disease identification method using machine learning classification in E-healthcare," *IEEE Access*, vol. 8, pp. 107562–107582, 2020, doi: 10.1109/ACCESS.2020.3001149.

[21] D. A. Noola and D. R. Basavaraju, "Corn leaf image classification based on machine learning techniques for accurate leaf disease detection," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 3, pp. 2509–2516, 2022, doi: 10.11591/ijece.v12i3.pp2509-2516.

[22] B. Shaheen and I. Németh, "Machine learning approach for degradation path prediction using different models and architectures of artificial neural networks," *Periodica Polytechnica Mechanical Engineering*, vol. 66, no. 3, pp. 244–252, 2022, doi: 10.3311/PPme.20145.

[23] A. S. Suryawanshi and N. Behera, "Prediction of abrasive wears behavior of dental composites using an artificial neural network," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 26, no. 6, pp. 710–720, 2023, doi: 10.1080/10255842.2022.2085509.

[24] M. Hamka and N. Ramdhoni, "K-means cluster optimization for potentiality student grouping using elbow method," in *AIP Conference Proceedings*, 2022, vol. 2578, no. 1, p. 60011, doi: 10.1063/5.0108926.

[25] G. Li, A. Zhang, Q. Zhang, D. Wu, and C. Zhan, "Pearson correlation coefficient-based performance enhancement of broad learning system for stock price prediction," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 69, no. 5, pp. 2413–2417, 2022, doi: 10.1109/TCSII.2022.3160266.

[26] H. Nguyen *et al.*, "Optimizing ANN models with PSO for predicting short building seismic response," *Engineering with Computers*, vol. 36, no. 3, pp. 823–837, 2020, doi: 10.1007/s00366-019-00733-0.

[27] A. Gehlot, B. K. Ansari, D. Arora, H. Anandaram, B. Singh, and J. L. Arias-Gonzáles, "Application of neural network in the prediction models of machine learning based design," in *Proceedings of the 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems, ICSES 2022*, 2022, pp. 1–6, doi: 10.1109/ICSES55317.2022.9914184.
[28] S. Kumar, M. D. Ansari, V. K. Gunjan, and V. K. Solanki, "On classification of BMD images using machine learning (ANN) algorithm," *Lecture Notes in Electrical Engineering*, vol. 601. pp. 1590–1599, 2020, doi: 10.1007/978-981-15-1420-3_165.
[29] L. C. Ribas, J. J. de M. Sá Junior, A. Manzanera, and O. M. Bruno, "Learning graph representation with Randomized Neural Network for dynamic texture classification," *Applied Soft Computing*, vol. 114, p. 108035, 2022, doi: 10.1016/j.asoc.2021.108035.
[30] I. S. Damanik, A. P. Windarto, A. Wanto, Poningsih, S. R. Andani, and W. Saputra, "Decision tree optimization in C4.5 algorithm using genetic algorithm," in *Journal of Physics: Conference Series*, 2019, vol. 1255, no. 1, p. 12012, doi: 10.1088/1742-6596/1255/1/012012.
[31] H. H. Patel and P. Prajapati, "Study and analysis of decision tree based classification algorithms," *International Journal of Computer Sciences and Engineering*, vol. 6, no. 10, pp. 74–78, 2018, doi: 10.26438/ijcse/v6i10.7478.
[32] Y. J. Mao, H. J. Lim, M. Ni, W. H. Yan, D. W. C. Wong, and J. C. W. Cheung, "Breast tumour classification using ultrasound Elastography with Machine Learning: A Systematic Scoping Review," *Cancers*, vol. 14, no. 2, p. 367, 2022, doi: 10.3390/cancers14020367.
[33] J. J. Tanimu, M. Hamada, M. Hassan, H. A. Kakudi, and J. O. Abiodun, "A machine learning method for classification of cervical cancer," *Electronics (Switzerland)*, vol. 11, no. 3, p. 463, 2022, doi: 10.3390/electronics11030463.
[34] A. Triayudi, W. O. Widyarto, L. Kamelia, Iksal, and Sumiati, "Clg clustering for dropout prediction using log-data clustering method," *IAES International Journal of Artificial Intelligence (IJAI)*, vol. 10, no. 3, pp. 764–770, 2021, doi: 10.11591/ijai.v10.i3.pp764-770.
[35] S. S. Yu, S. W. Chu, C. M. Wang, Y. K. Chan, and T. C. Chang, "Two improved k-means algorithms," *Applied Soft Computing Journal*, vol. 68, pp. 747–755, 2018, doi: 10.1016/j.asoc.2017.08.032.
[36] F. Duan, F. Song, S. Chen, M. Khayatnezhad, and N. Ghadimi, "Model parameters identification of the PEMFCs using an improved design of Crow Search Algorithm," *International Journal of Hydrogen Energy*, vol. 47, no. 79, pp. 33839–33849, 2022, doi: 10.1016/j.ijhydene.2022.07.251.
[37] B. J. D. Sitompul, O. Salim Sitompul, and P. Sihombing, "Enhancement clustering evaluation result of Davies-Bouldin index with determining initial centroid of K-means algorithm," in *Journal of Physics: Conference Series*, 2019, vol. 1235, no. 1, p. 12015, doi: 10.1088/1742-6596/1235/1/012015.
[38] R. Nainggolan, R. Perangin-Angin, E. Simarmata, and A. F. Tarigan, "Improved the performance of the K-means cluster using the sum of squared error (SSE) optimized by using the elbow method," in *Journal of Physics: Conference Series*, 2019, vol. 1361, no. 1, p. 12015, doi: 10.1088/1742-6596/1361/1/012015.
[39] H. Pan, X. You, S. Liu, and D. Zhang, "Pearson correlation coefficient-based pheromone refactoring mechanism for multi-colony ant colony optimization," *Applied Intelligence*, vol. 51, no. 2, pp. 752–774, 2021, doi: 10.1007/s10489-020-01841-x.
[40] P. Waldmann, "On the use of the pearson correlation coefficient for model evaluation in genome-wide prediction," *Frontiers in Genetics*, vol. 10, no. SEP, 2019, doi: 10.3389/fgene.2019.00899.
[41] J. Cai, M. H. Zhang, Y. T. Zhu, and Y. H. Liu, "Model of freight vehicle energy consumption based on pearson correlation coefficient," *Jiaotong Yunshu Xitong Gongcheng Yu Xinxi/Journal of Transportation Systems Engineering and Information Technology*, vol. 18, no. 5, pp. 241–246, 2018, doi: 10.16097/j.cnki.1009-6744.2018.05.035.
[42] H. Zhu, X. You, and S. Liu, "Multiple ant colony optimization based on pearson correlation coefficient," *IEEE Access*, vol. 7, pp. 61628–61638, 2019, doi: 10.1109/ACCESS.2019.2915673.
[43] T. Kogut and A. Slowik, "Classification of airborne laser bathymetry data using artificial neural networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1959–1966, 2021, doi: 10.1109/JSTARS.2021.3050799.
[44] G. Shobana and N. Priya, "Cancer drug classification using artificial neural network with feature selection," *Proceedings of the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV 2021*. pp. 1250–1255, 2021, doi: 10.1109/ICICV50876.2021.9388542.
[45] M. Yanto, S. Sanjaya, Yulasmi, D. Guswandi, and S. Arlis, "Implementation multiple linear regresion in neural network predict gold price," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 22, no. 3. Department of Informatics Engineering, Faculty of Computer Science, Universitas Putra Indonesia YPTK, Indonesia, pp. 1635–1642, 2021, doi: 10.11591/ijeecs.v22.i3.pp1635-1642.
[46] O. Adigun and B. Kosko, "Bidirectional Backpropagation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 5, pp. 1982–1994, 2020, doi: 10.1109/TSMC.2019.2916096.
[47] R. HECHT-NIELSEN, "Theory of the backpropagation neural network**based on 'nonindent' by Robert Hecht-Nielsen, which appeared in Proceedings of the International Joint Conference on Neural Networks 1, 593–611, June 1989. © 1989 IEEE.," in *Neural Networks for Perception*, 1992, pp. 65–93.
[48] M. O. Okwu and L. K. Tartibu, "Artificial neural network," in *Studies in Computational Intelligence*, vol. 927, 2021, pp. 133–145.
[49] H. Liu, "Optimal selection of control parameters for automatic machining based on BP neural network," *Energy Reports*, vol. 8, pp. 7016–7024, Nov. 2022, doi: 10.1016/j.egyr.2022.05.038.
[50] N. Jahan and R. Shahariar, "Predicting fertilizer treatment of maize using decision tree algorithm," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 3, pp. 1427–1434, 2020, doi: 10.11591/ijeecs.v20.i3.pp1427-1434.
[51] P. Sathiyanarayanan, S. Pavithra, M. Sai Saranya, and M. Makeswari, "Identification of breast cancer using the decision tree algorithm," in *2019 IEEE International Conference on System, Computation, Automation and Networking, ICSCAN 2019*, 2019, pp. 1–6, doi: 10.1109/ICSCAN.2019.8878757.
[52] D. Gustian and R. D. Hundayani, "Combination of AHP Method with C4.5 in the level classification level out students," in *3rd International Conference on Computing, Engineering, and Design, ICCED 2017*, 2018, vol. 2018-March, pp. 1–6, doi: 10.1109/CED.2017.8308098.

BIOGRAPHIES OF AUTHORS

Musli Yanto he is a lecturer at Universitas Putra Indonesia YPTK Padang. Educational background since 2014 has completed postgraduate studies in the field of informatics engineering. His skills include data science analysis, algorithms and programming, big data and artificial intelligence. He can be contacted at email: musli_yanto@upiyptk.ac.id.

Febri Hadi educational background comes from a Informatic of Engineering study program. Advanced education for the postgraduate program in Information System, Faculty of Computer Science, was completed in 2013. He is a lecturer at the Putra Indonesia University, YPTK Padang, majoring in information systems. Research history includes expert systems, decision support systems, and data mining. He can be contacted at email: febri_hadi@upiyptk.ac.id.

Syafri Arlis born in Padang on October 23, 1986. His undergraduate study was completed in 2009 at Universitas Putra Indonesia YPTK. He completed his Masters degree at Putra Indonesia University, YPTK Padang. He currently serves as a lecture in the Informatics Engineering study program at the Universitas Putra Indonesia YPTK Padang. Teaching history that has been carried out starting from 2011 until now, such as database and digital image processing. Published research history places more emphasis on the atificial neural network and digital image processing. He can be contacted at email: syafri_arlis@upiyptk.ac.id.