SISTEMASI: Jurnal Sistem Informasi
Volume 13, Nomor 1, Januari 2024: 167-176

ISSN:2302-8149
e-ISSN:2540-9719

Using Regression Model Analysis for Forecasting the Likelihood of Particular Symptoms of COVID-19

[1]**Agung Pangestu***, [2]**Ucu Sumirat**, [3]**Rosyid R. Al-Hakim**, [4]**Muhammad Yusro**, [5]**Risma Ekawati**, [6]**Mahmmoud H. A. Alrahman**, [7]**Machnun Arif**, [8]**Achmad Muchsin**, [9]**Nadhilla H. Wahyudiana**
[1]Electrical Engineering Department, Jakarta Global University, Depok, Indonesia
[2]Acyntia Teknologi Informasi Corporation, Depok, Indonesia
[3]Informatics Department, Universitas Harapan Bangsa, Purwokerto, Indonesia
[4]Electronic Engineering Department, State University of Jakarta, Jakarta, Indonesia
[5]Informatics Engineering Department, Jakarta Global University, Depok, Indonesia
[6]Department of Vertebrates, Iraqi Natural History Museum and Research Center, University of Baghdad, Baghdad, Iraq
[7]Master Program in Computer Science Department, Nusa Putra University, Sukabumi, Indonesia
[8]Independent Researcher, Jakarta, Indonesia
[9]Study Program of Natural Resources and Environmental Management Science (PSL), IPB University, Bogor, Indonesia
*e-mail: agungp@jgu.ac.id

(***received***: 2 Juli 2023, ***revised***: 27 Agustus 2023, ***accepted***: 7 November 2023)

Abstract

A certainty factor (CF) rule-based technique is frequently used by traditional expert systems (TES) in the medical industry to compute several symptoms and identify the inference solutions. The primary concern for this TES was predicting the likelihood of a particular ailment in the circumstances of new patients. Based on symptoms connected to clinical indicators in patients' diagnosis, CF is estimated. This TES probably won't be able to forecast unknown things, like the possibility of a particular ailment. Therefore, supervised learning techniques like linear regression can address this issue. We attempted to analyze the current COVID-19 TES by modeling the regression equation to forecast the chance of a particular disease that is COVID-like based on the CF value and the confidence level of the symptoms. To examine the most effective regression model to address the issue, we employed multi-linear regression (MLR) and multi-polynomial regression (MPR). The findings demonstrate that the MLR and MPR models are the most accurate regression models for estimating the chance of a disease associated with COVID-like symptoms. Our work built a basis for the creation of expert systems by concentrating more on MLES (machine learning expert systems) analytical techniques than TES.

Keywords: *expert system, forecasting, fuzzy, statistical linear regression, supervised learning algorithm.*

1 Introduction

The emergence of the novel coronavirus disease 2019 (COVID-19) caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus has ignited a global health crisis of unprecedented proportions [1]–[4]. As this pandemic continues to impact societies worldwide, effective management and mitigation strategies rely heavily on accurate and timely forecasting of disease spread and symptom likelihood [5], [6]. One of the crucial aspects in understanding and combating the spread of COVID-19 is the identification of specific symptoms that indicate the presence of the virus in individuals [7], [8].

In recent times, epidemiology and medical research have witnessed an accelerated utilization of data-driven approaches to comprehend and predict disease transmission dynamics [9], [10]. Leveraging the power of advanced statistical techniques and machine learning, researchers have made significant strides in modeling the spread of COVID-19 at a population level [11], [12]. However, there remains a compelling need to focus on predicting individual-level manifestations of the disease, especially the likelihood of experiencing particular symptoms associated with COVID-19.

Furthermore, the expert system tried to help medical experts for solving the patient's problem [13], [14]. Several studies about expert systems adopted in human medicine, including internal medicine

ISSN:2302-8149
e-ISSN:2540-9719

[15], [16], psychology [17]–[20], COVID-19 [21], [22], cancer [23]–[26], and traditional medicine [27], [28]. The expert system tried to perform human health problems, such as treatment, forecasting, and diagnosis [29]. Meanwhile, Al Hakim et al. [21] and Suprayitno et al. [30] developed the COVID-19 expert system based on mobile smartphones to solve the pandemic. Unfortunately, this current professional system has not been evaluated for performance and essential future forecasting purposes (for new patients with new symptoms). As a result, the study attempted to evaluate the performance of expert systems in regression modeling analysis using supervised learning (SL).

This research paper aims to contribute to the ongoing efforts in forecasting COVID-19 by employing regression model analysis to predict the probability of individuals exhibiting specific symptoms. By delving into a comprehensive dataset of confirmed COVID-19 cases and symptom profiles, we intend to construct predictive models that can offer insights into the likelihood of particular symptoms manifesting in infected individuals. The results of this study hold potential implications for public health interventions, clinical resource allocation, and personalized patient care. Therefore, this research is focusing on:

1. What are the results of the statistical analysis of the prediction of this study?
2. How to predict the specific likelihood of disease in symptoms of COVID-19 (COVID-like symptoms) using supervised machine learning based on the certainty factor method results?
3. If the prediction model is more effective, should it be applied to support new or related future symptoms of the COVID-19 variant?

This paper proposes to analyze a regression model for predicting (forecasting or estimating) COVID-19-related symptoms, i.e., the probability of a specific disease, such as symptoms similar to COVID-19. In a regression model, this study attempts to model the likelihood of particular COVID-19 symptoms in the future based on collected symptoms and improve expert system prediction performance. This paper supports the role of health information in expert system functions and is a new and timely paper, as machine learning (ML) is crucial to predicting future risks of COVID-19-related symptoms.

The document contributes to improving the performance of traditional expert systems using the method of certainty factors and provides an opportunity to use intelligent certainty factor calculation methods to predict the specific likelihood of diseases in cases of COVID-19. Health informatics, biomedical engineers, and computer scientists have a role in improving the performance of traditional systems methods of expertise, including safety factors.

2 Literature Review

The expert system was used to help some cases of human life [31], such as emergencies [32], [33], medical [21], [34], [35], engineering [36]–[38], education [39]–[41], business [42], [43], economics [44], military [45], [46], biological sciences [28], [36], [47]–[50], veterinary [51]–[54], agriculture [55]–[61], and forensics [62], [63]. As part of artificial intelligence, knowledge-based expert systems (ESs) are used to represent the knowledge of human experts in decision-making [64], [65].

Traditional expert systems (TES), such as the CF method, are used to calculate uncertainty, including knowledge based on human experts (CF rules) and users (CF users). One of the most popular algorithms used to calculate CF values is MYCIN [66]. This rule-based algorithm of TES tried calculating final CF values that represent symptoms (symptoms coded) [20], [21], [67], [68], or datasets of specific-related symptoms of some disease-causing [69]–[71]. The Equation (1) is the Equation of the CF value calculation [72].

$$CF = Rule_{CF} + Premise_{CF} \tag{1}$$

Based on Equation (1), the value of the symptom is from 0.00 (untrue) to 1.00 (true) for the certainty factors (CFs). The rule scheme also allows for the inclusion of conditional explanations of CFs. When a rule's premise is unclear due to uncertain symptoms, and the conclusion is uncertain due to the specification rule, the CF calculation containing the CF_{rule} ($Rule_{CF}$) and CF_{user} ($Premise_{CF}$) must be completed. The final CF value, $CF_{combine,}$ would determine the confidence level value [46].

Today, in the artificial intelligence (AI) era, the expert system was developed with the algorithm of machine learning [73], [74]. Supervised learning is one of the branches of machine learning, that is, supervised learning (SL) based on label information [75], and the algorithm of machine learning would be trained on it [74]. Besides, linear regression is the most common supervised learning method, a widely used and statistical-based procedure [76]. Supervised learning tries to solve any uncertain things [77] that are commonly the expert system's main issue, including medical issues [78]. Supervised learning methods operate on the assumption of building theories on existing data set instances to predict future data sets. A set of labelled cases is injected into a supervised learning algorithm that builds models for categorizing or anticipating future events [79].

Furthermore, research generally focuses on selecting the best regression model when analyzing regression methods and using regression equation models, which allows for determining the predictor or variable included in the regression process. The purpose of selecting the best regression model is usually in the interest of forecasting or prediction [80]. Furthermore, regression analysis types, including multi-linear regression (MLR) and multiplication polynomial regression (MPR), are usually used to model predictions based on existing data sets. MLR focused on the incorporation of each predictor and response variable. In addition, the MPR ignored the predictor and response variables' involvement [76].

As an intelligent tool to help experts solve problems, expert systems must be evaluated for performance, including prediction performance. Evaluating performance is necessary to predict future cases [79] based on knowledge-based representation datasets. Besides, the medical expert system is required to improve the system performance, including using regression analysis [79], [81].

The existing COVID-19 expert systems, developed by Al Hakim et al. [21] and Suprayitno et al. [30], are limited performance, especially in the case of particular disease likelihood of new symptoms and COVID-like symptoms. In addition, every confirmed patient has different physiological conditions, so they probably show different clinical signs. Besides, the existing expert system must evaluate the prediction performance for future forecasting cases. This study proposed two different regression analyses, multi-linear regression (MLR) and multiplication polynomial regression (MPR), for a better forecasting model. So, this study proposed the following hypotheses:

1. H_1: MLR regression model is the best forecasting model (if p-value < 0.05, one-tailed);
2. H_2: MPR regression model is the best forecasting model (if p-value < 0.05, one-tailed).

Both hypotheses might be acceptable to propose the best forecasting model; this study also tried to get the best-fit model based on R-squared, F-value, and p-value.

3 Research Method

Collecting, Acquisition, and Analysis of the Dataset

The study adopted COVID-19 symptoms data based on previous studies [21] (and to be developed in another mobile app version by Suprayitno et al. [30]). It authorized the identification of input functions (such as the confidence factor value (CF) based on symptoms collection and the confidence level (CL)) for decimal prediction of specific diseases based on symptoms information similar to COVID and the prediction of the impact of input functions and target variables. Input properties are studied using multi-linear regression (MLR) and multiplication polynomial regression (MPR), and this is one of the supervised machine learning algorithms used for regression tasks [29], [76], [82]. According to a related study [83], while MPR wouldn't contain any predictor variables for the CL variable (Y_2), MLR used CF_{rule} ($Rule_{CF}$) and CF_{user} ($Premise_{CF}$) as predictor variables (X_n) and confidence level (CL) as a response variable (Y_1). It is crucial to anticipate future CF values, especially in the case of new patients based on this dataset, and it was utilized for the best prediction of the value of the CF_{rule} and CF_{user} variables in conjunction with the CL variables that the regression equation can calculate. R Studio [84] was used in this study, including solving the MLR equation as seen in Equation (2), as well as the MPR equation that can be seen in the following Equation (3) [85].

$$Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \quad (2)$$

$$Y_2 = \beta_0 + \beta_1 X + \beta_2 X^2 + \cdots + \beta_k X^k + \varepsilon \quad (3)$$

within:

Y_1 = response variable (confidence level) of MLR;

 ISSN:2302-8149
e-ISSN:2540-9719

Y_2 = response variable (confidence level) of MPR;
β_0 = intercept;
$\beta_1, \beta_2, \dots, \beta_k$ = coefficients of regression;
$X_1, X_2, \dots, X_k$ = predictor variables (certainty factor values);
ε = disruptive factors that are unexplained by regression models.

Expert Knowledge Representation

A medical doctor (Aviasenna Andriand, MD) and supported medical references were adopted as knowledge-based representation datasets. Besides, a previous study by Al Hakim et al. [21] and improved by Suprayitno et al. [30] symptom datasets were also adopted for MYCIN rule-based algorithm calculation. To calculate the certainty factor value (CF), a rule-based method was utilized together with information from the user (diagnosed patient, then represented as $Premise_{CF}$ or CF_{user}) and expert (medical doctor, then represented as $Rule_{CF}$ or CF_{rule}). Each symptom was gathered in its section of the code. The dataset for this study was created by identifying each symptom as a code and determining that it had a CF between 0 and 1 decimally.

Modeling and Performance Evaluation

All gathered symptom CF values are labeled as COVID-like illness information. For its COVID-like illness label, this parameter (CF_{rule} and CF_{user}) used as predictor variables also generated the percentage of confidence level (CL, then utilized as a response variable). This dataset is extensively utilized to define the diagnosis's inference based on the inference machine employed, forward chaining, in the previous work by Al Hakim et al. [21] and Suprayitno et al. [30].

Al Hakim et al. [21] and Suprayitno et al. [30] categorized four COVID-like criteria: probable, suspect, positive, and negative. These studies also provided each criterion's confidence level (CL) %. Every ambiguity, including forecasts for COVID-like symptoms, must be made regularly. We attempted to utilize that dataset to predict continuous values based on a set of input attributes and to generate predictions on fresh data based on this dataset because such studies ([21], [30]) were unfortunately not mentioned in this topic. Both the statistical analysis of multiple linear regression (MLR) and multiple polynomial regression (MPR) findings will be examined in this study.

4 Results and Analysis

Multiple Linear Regression Result

According to the results of the MLR analysis, the F-value is 188.133, the p-value is 0.00 < 0.05 (one-tailed), and the significance is also shown (Table 1 for statistical reports). Equation (4) displays the regression model used in this regression study. This result demonstrates that the constant value of -0.2125 occurs when the CL variable (Y_1) is positive and the CF_{rule} predictor has not been impacted by the CF_{user} predictor or is in a constant state. Additionally, the CF_{rule} value (regression coefficient value) is 0.7116, which means that if it rises while other predictors, such as the CF_{user}, remain constant, the CL variable (Y_1) will rise. Additionally, the CF_{user} value (regression coefficient value) of 0.3659 indicates that the CL variable (Y_1) would rise if the CF_{user} value increased while the other variable (in this example, the CF_{rule} value) remained constant.

Table 1. The report of MLR analysis.

No.	Statistical Parameter	Result
1.	R-squared (R^2)	0.947
2.	R-value	0.973
3.	R-squared adjusted (adj. R^2)	0.942
4.	F-value	188.133
5.	p-value	0.00*

*Note: p-value < 0.05, one-tailed.

$$Y_1 = -0.2125 + 0.7116 \times CF_{rule} + 0.3659 \times CF_{user} \quad (4)$$

 ISSN:2302-8149
e-ISSN:2540-9719

Based on Equation 4, the degree of variability of the predictors (CF_{rule} and CF_{user}), which has been adjusted by the R-squared (coefficient of determination) value's weakness, is related to explaining its response variable (Y_1, which denotes CL for the MLR regression analysis), according to the R-squared adjusted value. The CL value is affected significantly by the combination of the CF variables, or this value is generally regarded as having a solid effect size, as indicated by the R-squared adjusted result of 0.942, which shows that CF_{rule} and CF_{user} have a 94.2% influence contributing to the CL value [86]. The remaining 5.8% of additional factors were not investigated in this study, but we made the assumption that they were the result of including each COVID-like symptom related to the conclusion of the CF calculation ($CF_{combine}$); this implies that the certainty factor method must be used to compare the CF_{user} and CF_{rule} to calculate the $CF_{combine}$, or final CF value [29]. This study performs a linear regression model related to Al-Hakim and Andriand [87], which studies leptospirosis cases using only one linear regression model. Besides, Al-Hakim et al. [83] revealed findings comparable to those of this study, namely that the multiple linear regression (MLR) was effectively described by the linear regression model, allowing the model to be used to predict illness in the future.

Multiple Polynomial Regression Result

The F-value for the MPR analysis result is 11,239.334, and its significance is shown by the p-value of 0.00 < 0.05 (one-tailed) (see Table 2 for statistical reports). Equation (5) displays the regression model used in this regression study. With a one-tailed p-value of less than 0.05, this result demonstrates the significance of all regression coefficients. The coefficient of determination (R-squared) is 100%, and the R-squared adjusted value is 99.9%. According to Moore et al. [86], a substantial effect size is deemed to exist if the R-squared value exceeds 0.7 (70%). The suggested MPR regression model is the best in this investigation since it outperforms the MLR model.

Table 2. The report of MPR analysis.

No.	Statistical Parameter	Result
1.	R-squared (R^2)	1.000
2.	R-value	1.000
3.	R-squared adjusted	0.999
4.	F-value	11,239.334
5.	p-value	0.00*

*Note: p-value < 0.05, one-tailed.

$$Y_2 = -0.0019 - 0.0858 \times CF_{rule} + 0.5244 \times {CF_{rule}}^2 + 0.0694 \times CF_{user} + 0.4897 \times {CF_{user}}^2 \quad (5)$$

The MPR regression model (Equation 5) shows that the predictors (CF_{rule} and CF_{user}) are not involved with each other, which means that both the patient's symptoms (source of CF_{user}) after the doctor's confirmed symptoms (source of CF_{rule}), are not related. This is important because each patient has a different body physiology condition, so it does not mean that the symptoms experienced by the patient always lead to the certainty of the doctor's anamnesis results. When this model (Equation 5) is used in new patients with new symptoms, it will better predict possible thyroid disorders in the future.

5 Conclusion and Recommendation

Based on the analysis of the two regression models, MLR (multi-linear or multiple linear regression) and MPR (multi-polynomial or multiple polynomial regression) can be applied to prediction models (both the p-values are lower than 0.05). However, MPR regression models provide better prediction capabilities for future cases of new patients, especially with symptoms similar to COVID-19. It is a foundation for developing an expert system focusing more on machine learning analysis methods (ML) than traditional rule-based expert systems.

However, our analysis does not guarantee the accuracy of the prediction of the existing expert system. It can only be used through statistical regression approaches to test which regression models are best for future predictions; of course, learning-assisted analysis is an alternative to expert system

 ISSN:2302-8149
e-ISSN:2540-9719

research, and hopefully, it will form the basis for further research related to improving the performance of expert systems.

Acknowledgement

We declare that there is no conflict of interest. The Research Management Centre (RMC) of the Jakarta Global University (JGU) supports the research. We are waiting for feedback from the study.

Reference

[1] A. A. Wani, M. Kumar, and I. Ahmed, "COVID-19 Structure Of Virus, Progression Of Disease & Present Status: A Review," *Int. J. Curr. Adv. Res.*, vol. 9, no. 5, pp. 22040–22043, 2020, doi: 10.24327/ijcar.2020.22043.4343.

[2] H. Sharma, "COVID-19: Public Health Emergency Of International Concern To Pandemic," *Int. J. Curr. Adv. Res.*, vol. 9, no. 06, pp. 22573–22574, 2020, doi: 10.24327/ijcar.2020.22574.4458.

[3] R. Lu *et al.*, "Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding," *Lancet*, vol. 395, no. 10224, pp. 565–574, Feb. 2020, doi: 10.1016/S0140-6736(20)30251-8.

[4] P. Zhou *et al.*, "A pneumonia outbreak associated with a new coronavirus of probable bat origin," *Nature*, vol. 579, no. 7798, pp. 270–273, 2020, doi: 10.1038/s41586-020-2012-7.

[5] R. Djalante *et al.*, "Review and analysis of current responses to COVID-19 in Indonesia: Period of January to March 2020," *Prog. Disaster Sci.*, vol. 6, p. 100091, Apr. 2020, doi: 10.1016/J.PDISAS.2020.100091.

[6] M. Palacios Cruz, E. Santos, M. A. Velázquez Cervantes, and M. León Juárez, "COVID-19, a worldwide public health emergency," *Rev. Clínica Española (English Ed.*, vol. 221, no. 1, pp. 55–61, 2021, doi: 10.1016/j.rceng.2020.03.001.

[7] A. Tahamtan and A. Ardebili, "Real-time RT-PCR in COVID-19 detection: issues affecting the results," *Expert Rev. Mol. Diagn.*, vol. 20, no. 5, pp. 453–454, 2020, doi: 10.1080/14737159.2020.1757437.

[8] A. E. Lim Jr., "Coronaviruses and Detection for SARS-CoV-2 Virus Causing COVID-19 Infection," *Innov. Technol. Manag. J.*, vol. 3, pp. 21–31, 2020.

[9] M. Naseem, R. Akhund, H. Arshad, and M. T. Ibrahim, "Exploring the Potential of Artificial Intelligence and Machine Learning to Combat COVID-19 and Existing Opportunities for LMIC: A Scoping Review," *J. Prim. Care Community Heal.*, vol. 11, 2020, doi: 10.1177/2150132720963634.

[10] S. Nurmaini, "The Artificial Intelligence Readiness for Pandemic Outbreak COVID-19: Case of Limitations and Challenges in Indonesia," *Comput. Eng. Appl.*, vol. 10, no. 1, pp. 9–19, Feb. 2021, doi: 10.18495/COMENGAPP.V10I1.353.

[11] W. E. Allen *et al.*, "Population-scale longitudinal mapping of COVID-19 symptoms, behaviour and testing," *Nat. Hum. Behav.*, vol. 4, no. 9, pp. 972–982, 2020, doi: 10.1038/s41562-020-00944-2.

[12] S. Saidi, S. Saidi, N. Herawati, and K. Nisa, "Modeling with generalized linear model on covid-19: Cases in Indonesia," *Int. J. Electron. Commun. Syst.*, vol. 1, no. 1, pp. 25–32, Jun. 2021, doi: 10.24042/ijecs.v1i1.9299.

[13] A. K. Meena and S. Kumar, "Study and Analysis of MYCIN expert system," *Int. J. Eng. Comput. Sci.*, vol. 4, no. 10, pp. 14861–14865, Oct. 2015, doi: 10.18535/ijecs/v4i10.41.

[14] W. Wiriyasuttiwong and W. Narkbuakaew, "Medical Knowledge-Based System for Diagnosis from Symptoms and Signs," *Int. J. Appl. Biomed. Eng.*, vol. 2, no. 1, pp. 54–59, 2009.

[15] M. Abdar, M. Zomorodi-Moghadam, R. Das, and I. H. Ting, "Performance analysis of classification algorithms on early detection of liver disease," *Expert Syst. Appl.*, vol. 67, pp. 239–251, 2017, doi: 10.1016/j.eswa.2016.08.065.

[16] I. Z. Jibril, J. Agajo, L. A. Ajao, J. G. Kolo, and O. C. Inalegwu, "Development of a Medical Expert System for Hypertensive Patients Diagnosis: A Knowledge-Based Rules," *Adv. Electr. Telecommun. Eng.*, vol. 1, no. 1, pp. 39–47, 2018.

[17] Y. Sha, T. Feng, X. Xiong, and T. Yang, "Designing Online Psychological Consultation Expert System Using Human-Computer Interaction," *Mob. Inf. Syst.*, vol. 2021, Jun. 2021, doi:

ISSN:2302-8149
e-ISSN:2540-9719

10.1155/2021/6458924.

[18] A. A. Al-Hajji, F. M. AlSuhaibani, and N. S. AlHarbi, "An Online Expert System for Psychiatric Diagnosis," *Int. J. Artif. Intell. Appl.*, vol. 10, no. 2, pp. 59–76, 2019, doi: 10.5121/ijaia.2019.10206.

[19] I. A. Alshawwa, M. Elkahlout, H. Q. El-Mashharawi, and S. S. Abu-Naser, "An Expert System for Depression Diagnosis," *Int. J. Acad. Heal. Med. Res.*, vol. 3, no. 4, pp. 20–27, 2019.

[20] N. K. Ariasih, "Expert System to Diagnose Diseases of Mental Health with Forward Chaining and Certainty Factor," *Wahana Mat. dan Sains J. Mat. Sains, dan Pembelajarannya*, vol. 14, no. 1, pp. 28–41, Apr. 2020, doi: 10.23887/WMS.V14I1.24267.

[21] R. R. Al Hakim, E. Rusdi, and M. A. Setiawan, "Android Based Expert System Application for Diagnose COVID-19 Disease: Cases Study of Banyumas Regency," *J. Intell. Comput. Heal. Informatics*, vol. 1, no. 2, pp. 1–13, 2020, doi: 10.26714/jichi.v1i2.5958.

[22] M. R. Mufid, A. Basofi, S. Mawaddah, K. Khotimah, and N. Fuad, "Risk diagnosis and mitigation system of covid-19 using expert system and web scraping," in *2020 International Electronics Symposium (IES)*, Surabaya (ID): Institute of Electrical and Electronics Engineers Inc., Sep. 2020, pp. 577–583. doi: 10.1109/IES50839.2020.9231619.

[23] G. Yu, Z. Chen, J. Wu, and Y. Tan, "Medical decision support system for cancer treatment in precision medicine in developing countries," *Expert Syst. Appl.*, vol. 186, p. 115725, Dec. 2021, doi: 10.1016/J.ESWA.2021.115725.

[24] R. Islam, S. Imran, M. Ashikuzzaman, and M. M. A. Khan, "Detection and Classification of Brain Tumor Based on Multilevel Segmentation with Convolutional Neural Network," *J. Biomed. Sci. Eng.*, vol. 13, no. 04, pp. 45–53, 2020, doi: 10.4236/jbise.2020.134004.

[25] V. V. Vadhiraj, A. Simpkin, J. O'Connell, N. S. Ospina, S. Maraka, and D. T. O'Keeffe, "Ultrasound Image Classification of Thyroid Nodules Using Machine Learning Techniques," *Medicina (Kaunas).*, vol. 57, no. 6, 2021, doi: 10.3390/medicina57060527.

[26] F. L. D. Cahyanti, W. Gata, and F. Sarasati, "Implementasi Algoritma Naïve Bayes dan K-Nearest Neighbor Dalam Menentukan Tingkat Keberhasilan Immunotherapy Untuk Pengobatan Penyakit Kanker Kulit," *J. Ilm. Univ. Batanghari Jambi*, vol. 21, no. 1, pp. 259–262, Feb. 2021, doi: 10.33087/JIUBJ.V21I1.1189.

[27] M. Melina, E. K. Putra, W. Witanti, S. Sukrido, and V. A. Kusumaningtyas, "Design and Implementation of Multi Knowledge Base Expert System Using the SQL Inference Mechanism for Herbal Medicine," *J. Phys. Conf. Ser.*, vol. 1477, p. 22007, 2020, doi: 10.1088/1742-6596/1477/2/022007.

[28] R. R. Al Hakim, H. A. Hidayah, A. Pangestu, D. Nugraha, S. Faizah, and E. R. C. Putri, "Sistem Pakar Forward-Chaining Dengan Certainty Factor Dalam Pemanfaatan Tumbuhan Obat, Studi Kasus: Etnobotani Di Kabupaten Banyumas, Jawa Tengah," in *E-Prosiding Seminar Nasional Inovasi Teknologi Pertanian Berkelanjutan (INOPTAN)*, 2022, pp. 106–113.

[29] E. Rich and K. Knight, *Artificial Intelligence*, 2nd ed. New York: McGraw-Hill Education, 1991.

[30] S. D. A. Suprayitno, M. N. Zakaria, and A. W. Yulianto, "COVID-19 Disease Diagnosis Expert System with Certainty Factor Method using iOS-Based App," *J. Telecommun. Netw. (Jurnal Jar. Telekomun.*, vol. 12, no. 3, pp. 160–165, 2022, doi: 10.33795/jartel.v12i3.336.

[31] M. I. Fale and Y. G. Abdulsalam, "Dr. Flynxz – A First Aid Mamdani-Sugeno-type fuzzy expert system for differential symptoms-based diagnosis," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 4, pp. 1138–1149, 2022, doi: 10.1016/j.jksuci.2020.04.016.

[32] X. SUN, H. LIU, G. WU, and Y. ZHOU, "Training effectiveness evaluation of helicopter emergency relief based on virtual simulation," *Chinese J. Aeronaut.*, vol. 31, no. 10, pp. 2000–2012, Oct. 2018.

[33] R. R. Al Hakim *et al.*, "Design and Development Smart-iMbulance for Efficiency of Road Emergency Priorities," *J. Innov. Res. Knowl.*, vol. 1, no. 2, pp. 167–172, 2021, doi: 10.53625/jirk.v1i2.

[34] M. I. Pramanik, R. Y. K. Lau, M. A. K. Azad, M. S. Hossain, M. K. H. Chowdhury, and B. K. Karmaker, "Healthcare informatics and analytics in big data," *Expert Syst. Appl.*, vol. 152, Aug. 2020.

[35] O. C. Ungureanu *et al.*, "Telemedicine Software Application," *Int. J. Comput. Sci. Netw. Secur.*, vol. 21, no. 2, pp. 171–180, 2021, doi: 10.22937/IJCSNS.2021.21.2.1.19.

ISSN:2302-8149
e-ISSN:2540-9719

[36] R. R. Al Hakim, G. E. Setyowisnu, and A. Pangestu, "An Expert System Dataset for Checking the Potential for Administering a Covid-19 Vaccine in Indonesia: Forward- Chaining Inference Machine Approach," *J. Glob. Eng. Res. Sci.*, vol. 1, no. 1, pp. 1–4, 2022, doi: 10.56904/jgers.v1i1.3.

[37] A. Farizi, “Sistem Pakar Untuk Mendiagnosa Kerusakan Komputer Dengan Menggunakan Metode Forward Chaining,” *Edu Komputika J.*, vol. 1, no. 2, pp. 21–32, 2014.

[38] M. M. Arifin and Y. B. Utomo, “Sistem Pakar Untuk Diagnosa Kerusakan Hardware Komputer Menggunakan Jaringan Syaraf Tiruan,” Dec. 2021.

[39] L. Salekhova, A. Nurgaliev, R. Zaripova, and N. Khakimullina, "The Principles of Designing an Expert System in Teaching Mathematics," *Univers. J. Educ. Res.*, vol. 1, no. 2, pp. 42–47, Aug. 2013, doi: 10.13189/UJER.2013.010202.

[40] S. Supriyadi and D. A. Wiliyanto, "Prototype Expert System Application to Identify Specific Children Learning Disabilities in Inclusion Schools," *IJDS Indones. J. Disabil. Stud.*, vol. 8, no. 1, pp. 117–127, May 2021, doi: 10.21776/UB.IJDS.2021.008.01.09.

[41] R. Eskrootchi, M. Zavari, and M. Alibeyk, "Developing a fuzzy expert system to determine the levels of students' eHealth literacy," *Libr. Philos. Pract.*, vol. 2021, pp. 1–26, Jan. 2021.

[42] Y. L. Lai and A. Ishizaka, "The application of multi-criteria decision analysis methods into talent identification process: A social psychological perspective," *J. Bus. Res.*, vol. 109, pp. 637–647, Mar. 2020, doi: 10.1016/j.jbusres.2019.08.027.

[43] A. P. Hamid, R. R. Al Hakim, A. Sungkowo, T. Trikolas, H. Purnawan, and A. Jaenul, “Talent Management Employee Development by Using Certainty Factor Method of Expert System,” *ARRUS J. Eng. Technol.*, vol. 1, no. 1, pp. 33–39, 2021, doi: 10.35877/jetech568.

[44] E. I. H. Rahayu, S. Suhardoyo, and I. Iwan, “Umpan Balik Sistem Pakar Sebagai Penilaian Kinerja Karyawan Pada PT. Multistrada Arah Sarana, Tbk,” *BMAJ Bus. Manag. Anal. J.*, vol. 2, no. 1, pp. 52–67, Apr. 2019, doi: 10.24176/bmaj.v2i1.3211.

[45] M. F. Azis, *Belajar Sendiri Pemrograman Sistem Pakar*. Jakarta (ID): Elex Media Komputindo, 1994.

[46] S. Harmanto, *Pengantar Sistem Pakar*. Depok (ID): Universitas Gunadarma, 1994.

[47] R. R. Al-Hakim *et al.*, “Sistem Pakar untuk Diagnosis Penyakit Tiroid dengan Gejala Psikologis Beserta Pengobatan Etnobotaninya,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 7, pp. 1771–1778, 2022, doi: 10.25126/jtiik.2022976763.

[48] B. A. Crawford, J. C. Maerz, and C. T. Moore, "Expert-informed habitat suitability analysis for at-risk species assessment and conservation planning," *J. Fish Wildl. Manag.*, vol. 11, no. 1, pp. 130–150, Jun. 2020, doi: 10.3996/092019-JFWM-075.

[49] M. Di Febbraro *et al.*, "Expert-based and correlative models to map habitat quality: Which gives better support to conservation planning?," *Glob. Ecol. Conserv.*, vol. 16, Oct. 2018, doi: 10.1016/J.GECCO.2018.E00513.

[50] Y. Darnita and R. Toyib, “Klasifikasi Penentuan Manfaat Tanaman Obat Herbal Berbasis Rule Based Reasoning,” *SISTEMASI*, vol. 10, no. 1, pp. 82–95, Jan. 2021, doi: 10.32520/STMSI.V10I1.1090.

[51] R. R. Al Hakim, A. Pangestu, and A. Jaenul, “Penerapan Metode Certainty Factor dengan Tingkat Kepercayaan pada Sistem Pakar dalam Mendiagnosis Parasit pada Ikan,” *Djtechno J. Inf. Technol. Res.*, vol. 2, no. 1, pp. 29–37, 2021, doi: 10.46576/djtechno.v2i1.1254.

[52] Suharjito, Diana, Yulyanto, and A. Nugroho, "Mobile Expert System Using Fuzzy Tsukamoto for Diagnosing Cattle Disease," *Procedia Comput. Sci.*, vol. 116, pp. 27–36, 2017.

[53] R. Stefani, “Sistem Pakar Diagnosa Penyakit Pada Ikan Koi Menggunakan Metode Backward Chaining,” *J. Ris. RUMPUN ILMU HEWANI*, vol. 1, no. 2, pp. 16–30, Oct. 2022, Accessed: Nov. 22, 2022.

[54] Z. Hakim and R. Rizky, “Sistem Pakar Diagnosis Penyakit Ikan Mas Menggunakan Metode Certainty Factor Di UPT Balai Budidaya Ikan Air Tawar Dan Hias Kabupaten Pandeglang Banten,” *J. Tek. Inform. Unis*, vol. 7, no. 2, pp. 164–169, 2020, doi: 10.33592/jutis.v7i2.399.

[55] E. Sudaryanto and A. Suryanto, “Sistem Pakar Diagnosa Hama Dan Penyakit Tanaman Durian Dengan Metode Naive Bayes,” *Teodolita Media Komunkasi Ilm. di Bid. Tek.*, vol. 21, no. 1, Jun. 2020, doi: 10.53810/JT.V21I1.339.

[56] M. H. Ramadhan, G. Dewantoro, and F. D. Setiaji, “Rancang Bangun Sistem Pakar Pemantau

ISSN:2302-8149
e-ISSN:2540-9719

Kualitas Air Berbasis IoT Menggunakan Fuzzy Classifier,” *J. Tek. Elektro*, vol. 12, no. 2, pp. 47–56, Dec. 2020, doi: 10.15294/jte.v12i2.25351.

[57] R. Hariyanto and K. Sa’diyah, “Sistem Pakar Diagnosis Penyakit dan Hama Pada Tanaman Tebu Menggunakan Metode Certainty Factor,” *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 3, no. 1, pp. 1–4, 2018, doi: 10.31328/jointecs.v3i1.500.

[58] S. H. Al Ikhsan, “Pengembangan Sistem Pakar Agribisnis Cabai (*Capsicum Annuum* L.) Berbasis Android,” Institut Pertanian Bogor, 2012.

[59] Y. -, F. Marisa, and D. Purnomo, “Sistem Rekomendasi Distribusi Tetes Tebu Di UD. Lancar Menggunakan Metode Fuzzy Sugeno Berbasis Web,” *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 1, no. 1, pp. 6–9, 2016, doi: 10.31328/jointecs.v1i1.401.

[60] Z. Zhao, F. Hong, H. Huang, C. Liu, Y. Feng, and Z. Guo, "Short-term prediction of fishing effort distributions by discovering fishing chronology among trawlers based on VMS dataset," *Expert Syst. Appl.*, vol. 184, 2021, doi: 10.1016/j.eswa.2021.115512.

[61] P. Karuniawan, I. N. Farida, and J. Suhertian, “Implementasi Metode Certainty Factor Untuk Mengidentifikasi Penyakit Tanaman Kedelai Dan Padi,” *Nusant. Eng.*, vol. 4, no. 1, p. 9, Apr. 2021, doi: 10.29407/NOE.V4I1.15902.

[62] K. Sturk-Andreaggi, M. A. Peck, C. Boysen, P. Dekker, T. P. McMahon, and C. K. Marshall, "AQME: A forensic mitochondrial DNA analysis tool for next-generation sequencing data," *Forensic Sci. Int. Genet.*, vol. 31, pp. 189–197, 2017, doi: 10.1016/J.FSIGEN.2017.09.010.

[63] J. Mortera, A. P. Dawid, and S. L. Lauritzen, "Probabilistic expert systems for DNA mixture profiling," *Theor. Popul. Biol.*, vol. 63, no. 3, pp. 191–205, 2003, doi: 10.1016/S0040-5809(03)00006-6.

[64] R. Sharda, D. Delen, and E. Turban, *Analytics, data science, & artificial intelligence: systems for decision support*, 11th ed. Hoboken, New Jersey (US): Pearson Education, 2020.

[65] Suyanto, *Artificial Intelligence: Searching, Reasoning, Planning, dan Learning*. Bandung (ID): Informatika, 2014.

[66] E. Hariadha, D. Nugraha, R. R. Al Hakim, A. Pangestu, M. Yusro, and M. H. Satria, "Using Certainty Factor for Symptoms Diagnosis of Thyroid Disorders," in *2022 International Conference on ICT for Smart Society (ICISS)*, Bandung (ID): IEEE, 2022, pp. 01–05. doi: 10.1109/ICISS55894.2022.9915219.

[67] A. S. Yanuar, E. G. Wahyuni, and D. T. Wiyanti, "Certainty Factor Method for Neurological Disease Diagnosis Based on Symptoms," in *ICCSET*, Kudus, Indonesia, 2018, pp. 851–856. doi: 10.4108/eai.24-10-2018.2280500.

[68] L. Safira, B. Irawan, and C. Setiningsih, "Implementation of the Certainty Factor Method for Early Detection of Cirrhosis Based on Android," *J. Phys. Conf. Ser.*, vol. 1201, no. 1, p. 12053, 2019, doi: 10.1088/1742-6596/1201/1/012053.

[69] C. Slamet, B. Firmanda, M. A. Ramdhani, W. Darmalaksana, E. Enjang, and F. M. Kaffah, "Android-based expert system design for drug selection using certainty factor," *J. Phys. Conf. Ser.*, vol. 1280, p. 22018, 2019, doi: 10.1088/1742-6596/1280/2/022018.

[70] D. Sudrajat *et al.*, "Expert system application for identifying formalin and borax in foods using the certainty factor method," *Eurasian J. Anal. Chem.*, vol. 13, no. 6, pp. 321–325, 2018.

[71] Sumiati, H. Saragih, T. K. A. Rahman, and A. Triayudi, "Expert system for heart disease based on electrocardiogram data using certainty factor with multiple rule," *IAES Int. J. Artif. Intell.*, vol. 10, no. 1, pp. 43–50, Mar. 2021, doi: 10.11591/ijai.v10.i1.pp43-50.

[72] X. Huang *et al.*, "A Generic Knowledge Based Medical Diagnosis Expert System," in *The 23rd International Conference on Information Integration and Web Intelligence (iiWAS2021)*, Linz (AT): ACM, New York, NY, USA, 2021, pp. 1–7. doi: 10.1145/3487664.3487728.

[73] N. Farsad Layegh, R. Darvishzadeh, A. K. Skidmore, C. Persello, and N. Krüger, "Integrating Semi-Supervised Learning with an Expert System for Vegetation Cover Classification Using Sentinel-2 and RapidEye Data," *Remote Sens.*, vol. 14, no. 15, p. 3605, Jul. 2022, doi: 10.3390/RS14153605.

[74] E. Lasso, T. T. Thamada, C. A. A. Meira, and J. C. Corrales, "Expert system for coffee rust detection based on supervised learning and graph pattern matching," *Int. J. Metadata, Semant. Ontol.*, vol. 12, no. 1, pp. 19–27, 2017, doi: 10.1504/IJMSO.2017.087641.

[75] R. Board and L. Pitt, "Semi-supervised learning," *Mach. Learn.*, vol. 4, pp. 41–65, Oct. 1989,

ISSN:2302-8149
e-ISSN:2540-9719

doi: 10.1007/BF00114803.
[76] R. Primartha, *Algoritma Machine Learning*. Bandung (ID): Informatika, 2021.
[77] D. D. Lewis and J. Catlett, "Heterogeneous Uncertainty Sampling for Supervised Learning," in *Machine Learning Proceedings 1994*, W. W. Cohen and H. Hirsh, Eds., New Brunswick (NJ): Morgan Kaufmann, 1994, pp. 148–156. doi: 10.1016/B978-1-55860-335-6.50026-X.
[78] E. B. Baum and F. Wilczek, "Supervised Learning of Probability Distributions by Neural Networks," *Neural Inf. Process. Syst.*, pp. 52–61, 1988.
[79] A. I. Hajamydeen and R. A. A. Helmi, "Performance of Supervised Learning Algorithms on Multi-Variate Datasets," in *Machine Learning and Big Data: Concepts, Algorithms, Tools and Applications*, wiley, 2020, pp. 209–232. doi: 10.1002/9781119654834.CH8.
[80] F. K. Lembang, "Analisis Regresi Berganda dengan Metode Stepwise pada Data Hbat," *J. Barekeng*, vol. 5, no. 1, pp. 15–20, 2011.
[81] S. Borzouei, H. Mahjub, N. Sajadi, and M. Farhadian, "Diagnosing thyroid disorders: Comparison of logistic regression and neural network models," *J. Fam. Med. Prim. Care*, vol. 9, no. 3, p. 1476, 2020, doi: 10.4103/JFMPC.JFMPC_910_19.
[82] R. Kohavi, "The power of decision tables," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 912, pp. 174–189, 1995, doi: 10.1007/3-540-59286-5_57.
[83] R. R. Al-Hakim *et al.*, "Predict the thyroid abnormality particular disease likelihood of the symptoms' certainty factor value and its confidence level: A regression model analysis," *Sist. J. Sist. Inf.*, vol. 12, no. 2, pp. 415–424, 2023, doi: 10.32520/stmsi.v12i2.2542.
[84] R Core Team, "R: A language and environment for statistical computing [Computer software manual]," 2016
[85] J. S. Malensang, H. Komalig, and D. Hatidja, "Pengembangan Model Regresi Polinomial Berganda Pada Kasus Data Pemasaran," *J. Ilm. Sains*, vol. 12, no. 2, pp. 149–152, Jan. 2013, doi: 10.35799/JIS.12.2.2012.740.
[86] D. S. Moore, W. I. Notz, and M. A. Flinger, *The basic practice of statistics*, 6th ed. New York (US): W.H. Freeman and Company, 2013.
[87] R. R. Al-Hakim and A. Andriand, "Diagnosis Leptospirosis Menggunakan Sistem Pakar Metode Faktor Kepastian Beserta Prediksi Regresi Linear," *Pros. SAINTEK Sains dan Teknol.*, vol. 2, no. 1, pp. 256–261, 2023.