

Predict The Thyroid Abnormality Particular Disease Likelihood of The Symptoms' Certainty Factor Value and Its Confidence Level: A Regression Model Analysis

^{1,2}Rosyid Ridlo Al-Hakim*, ^{1,3}Yanuar Zulardiansyah Arief, ¹Agung Pangestu, ⁴Hexa Apriliana Hidayah, ¹Aditia Putra Hamid, ⁵Aviasenna Andriand, ^{6,7}Nur Fauzi Soelaiman, ⁸Machnun Arif, ⁹Mahmmoud Hussein A. Alrahman

¹Electrical Engineering Department, Jakarta Global University, Depok, Indonesia

²Primateology Graduate Study Program, Graduate School, IPB University, Bogor, Indonesia

³Electrical & Electronic Engineering Department, Universiti Malaysia Sarawak, Samarahan, Malaysia

⁴Faculty of Biology, Universitas Jenderal Soedirman, Purwokerto, Indonesia

⁵Faculty of Medicine, Universitas Jenderal Soedirman, Purwokerto, Indonesia

⁶Computer and Informatics Department, Jakarta State Polytechnic, Depok, Indonesia

⁷Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Malaysia

⁸School of Computer Science, Nusa Putra University, Sukabumi, Indonesia

⁹Department of Vertebrates, Iraqi Natural History Museum and Research Center, University of Baghdad, Baghdad, Iraq

*e-mail: rosyidridlo10@gmail.com

(received: 3 Januari 2023, revised: 19 Maret 2023, accepted: 14 April 2023)

Abstract

The traditional expert system (TES) in the medical field commonly uses a certainty factor (CF) rule-based algorithm that can be calculated several symptoms to determine the inference solutions. The main issue for this TES included a prediction for some particular disease likelihood in the cases of new patients. CF is calculated based on symptoms related to clinical signs in patients' diagnoses. For some reason, this TES probably won't predict uncertain things, such as particular disease likelihood of some diseases. So, supervised learning, such as linear regression, can solve this problem. We tried to analyse the existing TES for thyroid disorders due to modelling the regression equation to predict the thyroid abnormality particular disease likelihood, based on the symptoms' CF value and its confidence level. We used multiple linear regression (MLR) and multiple polynomial regression (MPR) to analyse the best regression model to solve the problem. The results show that the MPR model indicates the best regression model for predicting particular disease likelihood of thyroid abnormality, supported by R-squared 94.7%, R-squared adjusted 94.4%, F-value 265.925, and p-value < 0.05, which are higher than MLR model. Our study proposed a foundation for expert system development by focusing more on machine learning expert system (MLES) analysis approaches than TES.

Keywords: expert system, fuzzy, linear regression, prediction, supervised learning algorithm.

1 Introduction

Thyroid glands regulate the hormone condition and are a crucial body organ supporting quality of life [1]. Thyroid patients might have different symptoms and are reported several psychological issues caused by abnormal T4 hormone regulation [2]. Patients with this illness frequently ignore their actions [3], [4]. Besides, in the era of health informatics, the role of the expert system (ES) was used to diagnose this issue [5].

ES, as the product of artificial intelligence (AI), tried to integrate with the medical expert for developing this system [6], [7]. Related studies about ES implemented in the medical field, such as COVID-19 [8], [9], psychology [10]–[13], traditional medicine [14], [15], internal medicine [16], [17], cancer [18]–[21]. ES is purposed to help medical experts solve health problems, including diagnosis, prediction, and treatment [22]. Meanwhile, Hariadha et al. [23] developed the medical ES to solve thyroid disorders. Besides, Al-Hakim et al. [5] coded this current ES to an Android-based app

<http://sistemasi.ftik.unisi.ac.id>

within the inference of traditional medication. Unfortunately, this current ES has not been evaluated for its performance nor for essential future prediction purposes (for new patients). So this study tried to evaluate its ES performance with supervised learning (SL) regarding the regression model analysis.

This paper proposes an analysis of a regression model for predicting (predicting or estimating) thyroid abnormality, which is the probability of a specific disease. In a regression model, this study tries to model the likelihood of particular diseases of thyroid abnormality based on the collected symptoms and increase the Expert System (ES) performance. This paper is a new and timely paper that supports the role of health information in the function of an expert system, as machine learning (ML) is vital in predicting the particular risk of thyroid abnormalities in thyroid diseases. Therefore, the research focuses on the following research questions:

1. What is the performance based on the statistical analysis results for prediction in this study?
2. How to predict the particular disease likelihood in thyroid abnormality's symptoms, based on the certainty method results using a supervised machine learning method?
3. If the prediction model better performing, should it be applied to support the future new or related symptoms of thyroid abnormality?

The paper contributes to performance upgrading in the traditional expert system (ES), which uses certainty factor method-based, providing the opportunity for predicting the particular disease likelihood in thyroid abnormality cases using an intelligent computing method. Health informaticists, biomedical engineers, and computer scientists put a role in upgrading the performance of the traditional expert system (ES) method, including the certainty factor method.

2 Literature Review

An expert system (ES) was used in daily human life, such as health [8], [24], [25], emergencies [26], [27], education [28]–[30], military [31], [32], economics [33], business management [34], [35], engineering [36]–[38], aeronautics [39], agriculture [40]–[46], biological sciences [5], [15], [36], [47]–[49], forensics [50], [51], and veterinary [52]–[55]. As part of artificial intelligence (AI), the expert system (ES) was used knowledge-based to represent the human expert knowledge in decision-making [56], [57]. An expert system was defined to help human life and assist the human expert in solving the problem [58].

The traditional expert system (TES), such as the certainty factor (CF) method, was used to calculate the uncertainty things that need to be transformed into certainty value, as well as containing knowledge based on human experts (CF_{rule}) and users (CF_{user}). One of the most rule-based algorithms used to calculate CF values is MYCIN [23]. This rule-based algorithm tried calculating CF values representing symptoms [8], [13], [59], [60], or specific-related symptom datasets [61]–[63]. The following equations (1) explain the calculation of the CF value [64].

$$CF = Rule_{CF} + Premise_{CF} \quad (1)$$

Based on equation (1), integers from 0 (for categorically untrue) to 1 make up the certainty factors (CFs) (for definitely true). The rule structure also allows for the inclusion of conditional statements of CFs. When a rule's premise is unclear owing to uncertain facts, and the conclusion is uncertain due to the rule's specification, the calculation of CF, including CF_{rule} ($Rule_{CF}$) and CF_{user} ($Premise_{CF}$), is performed. The final CF value determines the confidence level value called by $CF_{combine}$ [32].

In the era of artificial intelligence (AI), ES was developed and integrated with machine learning algorithms (MLAs) [65], [66]. One of the MLAs' algorithms is supervised learning (SL) based on the label information [67], and MLAs would be trained on it [66]. Besides, most SL analysis is linear regression, which is easily used and statistically procedurally [68]. SL tried to solve the uncertain things [69] commonly as the ES's central issue, including in the context of the medical problem [70]. SL methods work on the premise of constructing theories on existing dataset instances to predict future data instances. A set of labelled cases is fed into an SL algorithm, which builds a model to categorise or predict future occurrences [71].

Especially for regression analysis methods, usually for regression equation models, generally, research is focused on choosing the best regression model, where this procedure allows determining

the predictors or variables to be included in the regression. The purpose of selecting the best regression model is usually in the interest of forecasting or prediction [72]. Besides, types of regression analysis, including multiple linear regression (MLR) and multiple polynomial regression (MPR), are commonly used to model the prediction based on existing datasets. MLR focused on involving each predictor and response variable. Besides, MPR ignored the involvement of predictor and response variables [68].

As intelligent tools for helping an expert solve the problem, ES must be evaluated for performance, including prediction performance. Evaluating the performance was required to predict future instances [71] based on dataset knowledge-based representations. Besides, in the context of medical ES must investigate the performance, one of the best ways such as regression analysis [71], [73].

ES was developed to diagnose thyroid disorders [23] and several psychological signs for thyroid patients [5]. There are many implemented ES for thyroid cases [74]. For the gold standard issue, the thyroid must be required laboratory test results for its diagnosis [73]. Besides, each patient has a different physiological condition. In addition, Hariadha et al. [23] reported that thyroid symptoms, as well as clinical signs, may be different for each person.

Furthermore, as the conclusion of this section, as well as for future prediction reasons, existing ES must evaluate the performance of predictions. We proposed different regression analyses like MLR and MPR to get the best prediction model. So, this study proposed two hypotheses to support this issue:

1. H_1 : MLR regression model can be used for the best prediction model (if p-value < 0.05, one-tailed);
2. H_2 : MPR regression model can be used for the best prediction model (if p-value < 0.05, one-tailed).

These hypotheses might be acceptable to the prediction model; indeed, we tried to propose the best and good model based on R-squared, R-squared adjusted F-value, as well as p-value.

3 Research Method

Data Collection, Acquisition, and Analysis

Figure 1 shows the flowchart of this study. We adopted the thyroid dataset based on a previous study [5], [23] with permission to investigate the role of input features, such as certainty factor value (CF) based on symptom collection and confidence level (CL) in decimals, for the prediction of the likelihood of a particular disease based on thyroid disease information and modelling the relationship between the input features and the target variables. Investigating the input features using multiple linear regression (MLR) and multiple polynomial regression (MPR) analysis, as well as this one of the supervised machine learning algorithms used for regression tasks [22], [68], [75]. MLR involved CF_{rule} ($Rule_{CF}$) and CF_{user} ($Premise_{CF}$) as predictor variables (X_n) as well as confidence level (CL) as a response variable (Y_1), while the MPR wouldn't be involved between any predictor variables for the CL variable (Y_2). It was used for the best prediction of the value of the CF_{rule} and CF_{user} variables in conjunction with the CL variables that can be determined through the regression equation, and it is essential to predict future CF values, especially in the case of new patients based on this dataset. We used R Studio [76] to analyse the MLR problem, as seen in equation (2), as well as the MPR equation can be seen in the following equation (3) [77], included in this study.

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

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

within:

- Y_1 = response variable (CL) for the MLR regression analysis;
- Y_2 = response variable (CL) for the MPR regression analysis;
- β_0 = intercept;
- $\beta_1, \beta_2, \dots, \beta_k$ = regression coefficients;
- X_1, X_2, \dots, X_k = predictor variables (CF);

ϵ = disruptive factors that cannot be explained by regression models.

Expert Interview

Symptom collection was adopted from a medical doctor (Aviasenna Andriand, MD) and supported medical references, including New Castle Index (NCI), Wayne Index (WI), and Billewicz Index (BI). Based on a previous study by Hariadha et al. [23], as well as improved by Al-Hakim et al. [5], this symptom dataset was also used for MYCIN rule-based algorithm calculation. A rule-based algorithm was used for the basis of certainty factor value (CF) calculation, as well as adopted knowledge-based from the expert (MD, then represented as Rule_{CF} or CF_{rule}) and user (diagnosed patient, then represented as Premise_{CF} or CF_{user}). These 64 symptoms were collected within the code, respectively. Each symptom was identified as a code as well as containing a CF between 0 and 1 decimally, and then it was used as the dataset for this study.

Data Modelling and Evaluation

CF values for 64 symptoms are categorised as thyroid disease information (label). This parameter (CF_{rule} and CF_{user}) used as predictor variables also calculated the percentage of confidence level (CL, then used as a response variable) for its thyroid disease label, respectively. According to the previous study by Hariadha et al. [23] and Al-Hakim et al. [5], this dataset is significantly used to determine the diagnosis's inference based on the inference machine used is forward chaining.

Hariadha et al. [23] and Al-Hakim et al. [5] categorised four thyroid diseases: hypothyroidism, hyperthyroidism, goitre, and thyroid nodule. These studies also reported the percentage of confidence level (CL) for each disease. Meanwhile, every uncertainty, including thyroid symptom predictions, must be made continuously. Unfortunately, those study ([5], [23]) was not discussed in this issue, so we tried to use that dataset to predict continuous values based on a set of input features and to make predictions on new data based on this dataset. The results that would be analysed in this paper include statistical analysis of multiple linear regression and statistical analysis of multiple polynomial regression.

4 Results and Analysis

Statistical Analysis of Multiple Linear Regression

The result of MLR analysis shows that the F-value is 3.394 and the p-value $0.04 < 0.05$ (one-tailed), as well as its significance (Table 1 for statistical reports). The regression model for this regression analysis shows in equation (4). This result shows that the constant value of 0.7439 is when the CF_{rule} predictor has not been affected by the CF_{user} predictor or is in a constant state, and then the CL variable (Y_1) is positive. Besides, the CF_{rule} value (regression coefficient value) is 0.16, meaning that if the CF_{rule} value increases while other predictors are constant, in this case, the CF_{user}, then the CL variable (Y_1) would also increase. In addition, the CF_{user} value (regression coefficient value) is 0.133 means that if the CF_{user} value increases while the other variable (in this case, it is the CF_{rule} value) is constant, then the CL variable (Y_1) would also increase.

Table 1. MLR analysis reports for this study

No.	Statistical Parameter	Result
1.	R-squared	0.100
2.	R-value	0.316
3.	R-squared adjusted	0.071
4.	F-value	3.394
5.	p-value	0.04

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

$$Y_1 = 0.7439 + 0.16 \times CF_{rule} + 0.133 \times CF_{user} \quad (4)$$

Based on the R-squared adjusted value, which determines how much the degree of variability of the predictors (CF_{rule} and CF_{user}) that has been adjusted by the weakness of the R-squared (coefficient of determination) value is related to explaining its response variable (Y_1 , which means CL for the

<http://sistemasi.ftik.unisi.ac.id>

MLR regression analysis). The R-squared adjusted result is 0.071, which means that CF_{rule} and CF_{user} have a 7.1% influence contributing to the CL value, so this indicates that the combination of CF variables affects the value of the CL variable in a small way or this value is generally considered a none or very weak effect size [78]. The remaining 92.9% of other factors were not studied in this study, but we assumed that it was caused by involving each symptom related to the end of CF calculation ($CF_{combine}$) [22]; it means that the certainty factor method must be involved the CF_{rule} against CF_{user} for the calculation of the $CF_{combine}$ or final CF value [22]. The study proposed a linear regression model, which is related to the research by Al-Hakim and Andriand [79] for leptospirosis cases, but only used a single linear regression model. In this study, the linear regression model has successfully modelled the multiple linear regression, so the model would be implemented for disease prediction.

Statistical Analysis of Multiple Polynomial Regression

Meanwhile, the MPR analysis result shows that the F-value is 265.925 and the p-value is $0.00 < 0.05$ (one-tailed), as well as its significance (Table 2 for statistical reports). The regression model for this regression analysis shows in equation (5). This result shows that all the regression coefficients are significant, supported by p-value < 0.05 , one-tailed, while the coefficient of determination (R-squared) is 94.7% and the R-squared adjusted is 94.4%. According to Moore et al. [78], if the R-squared value is more than 0.7, this is considered a large impact size. This MPR model is better than the MLR model, so the best regression model in this study is the proposed MPR regression model.

Table 2. MPR analysis reports for this study

No.	Statistical Parameter	Result
1.	R-squared	0.947
2.	R-value	0.973
3.	R-squared adjusted	0.944
4.	F-value	265.925
5.	p-value	0.00*

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

$$Y_2 = 2.3991 \times CF_{rule} - 1.875 \times CF_{rule}^2 + 0.9685 \times CF_{user} - 0.804 \times CF_{user}^2 \quad (5)$$

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 our analysis of the two regression models, MLR (multiple linear regression) and MPR (multiple polynomial regression) are acceptable to the prediction model (both show a significant value with p-value < 0.05). However, the MPR regression model provides better predictive capabilities for new patient cases in the future. It can be a foundation for expert system (ES) development by focusing more on machine learning (ML) analysis approaches than traditional rule-based expert systems.

However, our analysis does not provide certainty about the correctness of predictions from existing expert systems. It only helps through a statistical regression approach to test which regression model is better in terms of future predictions; of course, with this approach, learning-supervised analysis can be an alternative to the study of expert systems, and hopefully, it can be a foundation for further research related to improving the performance of expert systems.

Acknowledgement

We declare that it does not have any conflict of interest. We also acknowledged all researchers ([5], [23]) that helped us evaluate the existing expert system. Research Management Center (RMC) Jakarta Global University (JGU) Indonesia supports this study. We are eagerly awaiting any feedback on this study.

Reference

- [1] Azamris, *Buku Ajar Kelainan Tiroid*. Yogyakarta (ID): Deepublish Publisher, 2020.
- [2] A. Fukao *et al.*, “Graves’ disease and mental disorders,” *Journal of Clinical and Translational Endocrinology*, vol. 19. Elsevier, p. 100207, Mar. 01, 2020. doi: 10.1016/j.jcte.2019.100207.
- [3] L. L. Altshuler *et al.*, “Does thyroid supplementation accelerate tricyclic antidepressant response? A review and meta-analysis of the literature,” *Am. J. Psychiatry*, vol. 158, no. 10, pp. 1617–1622, 2001, doi: 10.1176/appi.ajp.158.10.1617.
- [4] M. Bauer, T. Goetz, T. Glenn, and P. C. Whybrow, “The thyroid-brain interaction in thyroid disorders and mood disorders,” *J. Neuroendocrinol.*, vol. 20, no. 10, pp. 1101–1114, 2008, doi: 10.1111/J.1365-2826.2008.01774.X.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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)*, Sep. 2020, pp. 577–583. doi: 10.1109/IES50839.2020.9231619.
- [10] 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: 10.1155/2021/6458924.
- [11] 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/ijai.2019.10206.
- [12] I. A. Alshawwa, M. Elkahout, 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.
- [13] 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.
- [14] 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.
- [15] 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 (An Expert System With Forward-Chaining And Certainty Factor For Medicinal Plant Uses, Cases Study: Ethnobotana),” in *E-Prosiding Seminar Nasional Inovasi Teknologi Pertanian Berkelanjutan (INOPTAN)*, 2022, vol. 1, no. 1, pp. 106–113. [Online]. Available: <https://jurnal.unikastpaulus.ac.id/index.php/inoptan/article/view/1412>
- [16] 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.

<http://sistemasi.ftik.unisi.ac.id>

- 239–251, 2017, doi: 10.1016/j.eswa.2016.08.065.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] E. Rich and K. Knight, *Artificial Intelligence*, 2nd ed. New York: McGraw-Hill Education, 1991.
- [23] 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)*, 2022, pp. 01–05. doi: 10.1109/ICISS55894.2022.9915219.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] R. R. Al Hakim *et al.*, “Design and Development Smart-iMbalance for Efficiency of Road Emergency Priorities,” *J. Innov. Res. Knowl.*, vol. 1, no. 2, pp. 167–172, 2021, doi: 10.53625/jirk.v1i2.
- [28] 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.
- [29] S. Supriyadi and D. Wiliyanto, “Prototype Expert System Application to Identify Specific Children Learning Disabilities in Inclusion Schools,” *IJDS Indones. J. Disabil. Stud.*, vol. 8, no. 01, pp. 117–127, Jan. 2021, doi: 10.21776/ub.ijds.2021.008.01.09.
- [30] 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.
- [31] M. F. Azis, *Belajar Sendiri Pemrograman Sistem Pakar*. Jakarta (ID): Elex Media Komputindo, 1994.
- [32] H. S. Suryadi, *Pengantar Sistem Pakar*. Depok: Universitas Gunadarma, 1994.
- [33] 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.
- [34] 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.
- [35] A. P. Hamid, R. R. Al Hakim, A. Sungkowo, T. Trikolos, 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.
- [36] R. R. Al Hakim, G. E. Setyowisnu, and A. Pangestu, “An Expert System Dataset for Checking <http://sistemasi.ftik.unisi.ac.id>

- 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. Accessed: Apr. 28, 2021. [Online]. Available: <https://ejournal.uniska-kediri.ac.id/index.php/JTECS/article/view/715>
- [39] A. Eresen, N. Imamoğlu, and M. Önder Efe, “Autonomous quadrotor flight with vision-based obstacle avoidance in virtual environment,” *Expert Syst. Appl.*, vol. 39, no. 1, pp. 894–905, Jan. 2012, doi: 10.1016/J.ESWA.2011.07.087.
- [40] E. Sudaryanto and A. Suryanto, “Sistem pakar diagnosa hama dan penyakit tanaman durian dengan metode naive bayes,” *Teodolita Media Komunikasi Ilm. di Bid. Tek.*, vol. 21, no. 1, Jun. 2020, doi: 10.53810/JT.V21I1.339.
- [41] M. H. Ramadhan, G. Dewantoro, and F. D. Setiaji, “Rancang Bangun Sistem Pakar Pemantau 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.
- [42] 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.
- [43] S. H. Al Ikhsan, “Pengembangan Sistem Pakar Agribisnis Cabai (*Capsicum Annuum L.*) Berbasis Android,” Institut Pertanian Bogor, 2012.
- [44] 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.
- [45] 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.
- [46] 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.
- [47] 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.
- [48] 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.
- [49] 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.
- [50] K. Sturk-Andreaggi, M. A. Peck, C. Boysen, P. Dekker, T. P. McMahan, 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.
- [51] 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.
- [52] 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.
- [53] 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.
- [54] 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. [Online]. Available: <https://prin.or.id/index.php/JURRIH/article/view/526>
- [55] Z. Hakim and R. Rizky, “Sistem Pakar Diagnosis Penyakit Ikan Mas Menggunakan Metode <http://sistemasi.ftik.unisi.ac.id>

- 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.
- [56] 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.
- [57] Suyanto, *Artificial Intelligence: Searching, Reasoning, Planning, dan Learning*. Bandung: Informatika, 2014.
- [58] M. I. Fale and Y. G. Abdulsalam, “Dr. Flynnx – 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.
- [59] A. S. Yanuar, E. G. Wahyuni, and D. T. Wiyanti, “Certainty Factor Method for Neurological Disease Diagnosis Based on Symptoms,” in *ICCSET*, 2018, pp. 851–856. doi: 10.4108/eai.24-10-2018.2280500.
- [60] 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.
- [61] 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.
- [62] 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.
- [63] 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.
- [64] X. Huang *et al.*, “A Generic Knowledge Based Medical Diagnosis Expert System,” in *The 23rd International Conference on Information Integration and Web Intelligence (iiWAS2021)*, 2021, pp. 1–7. doi: 10.1145/3487664.3487728.
- [65] 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.
- [66] 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.
- [67] R. Board and L. Pitt, “Semi-supervised learning,” *Mach. Learn.*, vol. 4, pp. 41–65, Oct. 1989, doi: 10.1007/BF00114803.
- [68] R. Primartha, *Algoritma Machine Learning*. Bandung (ID): Informatika, 2021.
- [69] 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.
- [70] E. B. Baum and F. Wilczek, “Supervised Learning of Probability Distributions by Neural Networks,” *Neural Inf. Process. Syst.*, pp. 52–61, 1988.
- [71] 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.
- [72] F. K. Lembang, “Analisis Regresi Berganda dengan Metode Stepwise pada Data Hbat,” *J. Barekeng*, vol. 5, no. 1, pp. 15–20, 2011.
- [73] 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.
- [74] R. R. Al Hakim, M. H. Satria, Y. Z. Arief, A. D. Setiawan, A. Pangestu, and H. A. Hidayah, “Artificial Intelligence for Thyroid Disorders: A Systematic Review,” *Sci. Inf. Technol. Lett.*, vol. 2, no. 2, Nov. 2021, doi: 10.31763/sitech.v2i2.694.
- [75] 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.

- [76] R Core Team, “R: A language and environment for statistical computing [Computer software manual],” 2016
- [77] 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.
- [78] D. S. Moore, W. I. Notz, and M. A. Fligner, *The basic practice of statistics*, 6th ed. New York (US): W.H. Freeman and Company, 2013.
- [79] 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, [Online]. Available: <https://www.jurnal.pelitabangsa.ac.id/index.php/SAINTEK/article/view/2165>