Pertanika J. Sci. & Technol. 31 (6): 2723 - 2733 (2023)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Development of Artificial Neural Network Model for Medical Specialty Recommendation

Winda Hasuki[1], David Agustriawan[1], Arli Aditya Parikesit[1]*, Muammar Sadrawi[1], Moch Firmansyah[2], Andreas Whisnu[3], Jacqulin Natasya[1], Ryan Mathew[4], Florensia Irena Napitupulu[5] and Nanda Rizqia Pradana Ratnasari[1]

[1]*Department of Bioinformatics, School of Life Sciences, Indonesia International Institute for Life Sciences, Jakarta 13210, Indonesia*
[2]*Information Technology Department, Lira Medika Hospital, Jawa Barat 41314, Indonesia*
[3]*Information Technology Department, Indonesia International Institute for Life Sciences, Jakarta 13210, Indonesia*
[4]*Department of Biotechnology, School of Life Sciences, Indonesia International Institute for Life Sciences, Jakarta 13210, Indonesia*
[5]*Department of Food Science, School of Life Sciences, Indonesia International Institute for Life Sciences, Jakarta 13210, Indonesia*

ABSTRACT

Timely diagnosis is crucial for a patient's future care and treatment. However, inadequate medical service or a global pandemic can limit physical contact between patients and healthcare providers. Combining the available healthcare data and artificial intelligence methods might offer solutions that can support both patients and healthcare providers. This study developed one of the artificial intelligence methods, artificial neural network (ANN), the multilayer perceptron (MLP), for medical specialist recommendation systems. The input of the system is symptoms and comorbidities. Meanwhile, the output is the medical specialist. Leave one out cross-validation technique was used. As a result, this study's F1 score of the model was about 0.84. In conclusion, the ANN system can be an alternative to the medical specialist recommendation system.

Keywords: Machine learning, medical specialty, multilayer perceptron, neural network, recommendation

ARTICLE INFO

Article history:
Received: 19 September 2022
Accepted: 25 January 2023
Published: 08 September 2023

DOI: https://doi.org/10.47836/pjst.31.6.05

E-mail addresses:
winda.hasuki@student.i3l.ac.id (Winda Hasuki)
david.agustrawan@i3l.ac.id (David Agustriawan)
arli@daad-alumni.de (Arli Aditya Parikesit)
muammar.sadrawi@i3l.ac.id (Muammar Sadrawi)
moch.firmansyah@pundiraya.co.id (Moch Firmansyah)
johnkecops@yahoo.co.uk (Andreas Whisnu)
jacqulin.natasya@student.i3l.ac.id (Jacqulin Natasya)
ryan.mathew@student.i3l.ac.id (Ryan Mathew)
florensia.napitupulu@i3l.ac.id (Florensia Irena Napitupulu)
nanda.ratnasari@i3l.ac.id (Nanda Rizqia Pradana Ratnasari)
* Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

INTRODUCTION

Digitalization of healthcare provides new possibilities to improve healthcare services. One problem that can arise in health care is limited physical contact between patients and healthcare providers, which can prolong the diagnosis process. Patients must rely on healthcare providers, such as primary care physicians, to determine the most suitable medical specialty for treating their condition. The delayed diagnosis can result in missed opportunities for intervention. Contact between patients and healthcare providers can be hindered due to various situations. For example, during the COVID-19 pandemic, it was harder for patients to meet healthcare providers (Lee et al., 2021). Such situations can also happen to patients who live in remote or impoverished areas.

With the increasing number of healthcare data, incorporating artificial intelligence (AI) methods can offer useful insights to help decision-making. One of the most popular methods in medical applications is the artificial neural network (ANN) (Jiang et al., 2017). The architecture of ANN consists of three types of layers: input, hidden, and output. Information or features from the external environment would be received by the input layer and passed to the hidden layers. In the hidden layers, the pattern of data would be extracted. After that, the output layer would present the network output (da Silva et al., 2017). In the simplest form, ANN can have no hidden layer, making it a single-layer neural network that consists of only input and output layers. Models with one or more hidden layers are usually referred to as multilayer neural networks. ANN can have a feed-forward or feedback architecture.

ANN has been applied in various studies, including classification, prediction, and diagnosis (Shahid et al., 2019). This method is chosen for many studies as it can process a large amount of data. It is also found to be less likely to overlook important information, and it can reduce diagnosis time (Amato et al., 2013). Although ANN is often applied for large datasets, some studies suggested that this method can also work for small datasets. Feng et al. (2019) created a model to predict material defects with 487 data. Their deep neural network model showed high accuracy. The ANN-based dengue predictive model also reached high accuracy, sensitivity, and specificity and was trained with a small dataset (Silitonga et al., 2021). In another study, a model trained with 116 data achieved an accuracy of 82% without signs of overfitting (Olson et al., 2018).

This study developed a classification model using ANN, focusing on one type of feed-forward neural network, which is the multilayer perceptron (MLP). MLP is fully connected, and it uses backpropagation during training. Some common machine-learning Python libraries were used to build, train, and test the MLP model, including Keras, TensorFlow, and scikit-learn. This study trained the model with a small dataset containing 111 data points. This dataset consisted of common symptoms, comorbidities, and medical specialties. Leave-one-out cross-validation (LOOCV) was also conducted to estimate the performance of the model. This study aimed to create a model that can recommend suitable medical specialties for patients based on their symptoms and comorbidities.

RELATED WORK

ANN in Healthcare

ANN has been implemented in many diagnostics models. Yao et al. (2019) applied ANN with a backpropagation procedure to diagnose diabetic retinopathy. This model was developed based on the results of multivariable logistic regression (MLR) using data from 530 residents. The AUC of the ANN model reached 0.84, which was higher than the AUC of MLR. Aguiar et al. (2016) developed two ANN models to support the diagnosis of pulmonary tuberculosis. The first was an MLP model for classification, while the second was for risk assignment. Both models showed good performance. In another study, an MLP model was developed to classify hypertension (Chai et al., 2021). The model achieved an accuracy of 0.76 and an AUC of 0.75. They noted that although their model cannot be used as a clinical decision-making tool, it can still be used as an early warning mechanism.

Besides diagnosis, ANN has also been used to help decision-making management. The ANN model was applied for syncope risk stratification in the Casagranda et al. (2016) study. This study also compared the ANN model to a multivariate logistic regression model. The ANN model showed better performance. Ippoliti et al. (2021) proposed ANN models to predict the outcome of hospital admission. The results showed that applying their models could reduce the average length of stay. The result of these studies can be useful for hospital managers. On a global scale, a forecasting model for the COVID-19 time series was created with ANN (Borghi et al., 2021). Their study included data from 30 countries. The model they developed could predict time series behavior related to the number of COVID-19 infections and deaths. ANN has also been applied to analyze corruption in healthcare (Buscema et al., 2017). The result of this study suggested that countries with similar Human Development Index would perceive corruption in the same way.

AI Methods for Medical Specialty Recommendation

Various algorithms have been applied to create a model for disease prediction or doctor recommendations for different purposes. Rémy et al. (2018) created a logistic regression model to predict which healthcare practitioner is the most appropriate for a patient. Some diseases may require multiple specialists to provide an accurate diagnosis, and their study aimed to see if it is possible to form a group of physicians using a probabilistic model. Their proposed model also assumed that 4 symptoms were enough to describe a disease and that patients have no comorbidities. A study by Kumar et al. (2021) also tested four machine learning approaches—Naive Bayes, random forest, logistic regression, and KNN. They created a website where users enter their symptoms and get medical specialty recommendations (pro-prediction). They also provide an option for specific predictions, such as prediction for heart disease and diabetes. The dataset that they used for pro-prediction was coded into binary classification. The four models showed good accuracy for pro-prediction, with random forest showing the best performance (90.2%). In another study, Lee et al. (2021) applied deep learning-based NLP. They developed an AI chatbot that can give medical specialty recommendations. Their dataset consisted of 51,134 sentences and included data for 26 medical specialties. Implementing MLP in medical specialty recommendations has not been found in any literature but has been used for other applications such as diagnosis models.

Cross-validation Methods for Small Dataset

In the development of machine learning models, cross-validation is often applied to evaluate the performance of the models. There are many cross-validation methods. The simplest and most common method is the hold-out cross-validation, in which the dataset is divided into training and testing sets, and the model would usually be trained only once. Usually, 80% of the dataset is used as the training set, while the remaining 20% is used for the testing set. However, this method is more suitable for large datasets. Another common method is k-fold cross-validation. For this method, there would be k iterations of training and validation; for each iteration, different segments would be used for validation (Refaeilzadeh et al., 2009). There is also the leave-one-out cross-validation (LOOCV), where the number of segments would be equal to the number of instances in the dataset (Webb et al., 2011). This method shows a lower bias compared to k-fold. However, it also shows higher variability in test error. In larger datasets, LOOCV is considered computationally expensive, but in small datasets, it can maximize the extension of the training set (Pasini, 2015).

Problem Statement

Some situations can lead to patients being unable to see primary care providers for a diagnosis and referral to medical specialists. Models proposed in other studies often ignored comorbidities, even though it is important in deciding which medical specialty suits a patient's condition. This study addressed this problem by developing a model to give medical specialty recommendations based on the symptoms and comorbidities.

METHODOLOGY

This study proposed an MLP model for a medical specialty recommendation system. The workflow consisted of four steps: data collection and cleaning, model building, model training and testing, and performance evaluation. All steps in this study were done in Python 3.6.13, with some libraries including Pandas, Numpy, TensorFlow, Keras, and scikit-learn. For the first step, Pandas and Numpy were used. Data on common symptoms and comorbidities, as well as the suitable medical specialties, were collected, creating a dataset for a multi-class classification problem. The categorical variables were coded into binary classifications (0 or 1). There were some missing and ambiguous values found. These data were removed from the dataset. After that, the dataset was separated into input and output data. The output data was then converted into an array.

The MLP model was created by using Keras and Tensorflow. MLP was chosen as it is suitable for our dataset, which has labeled input. Due to the size of the dataset, it is not necessary to apply a more advanced model as well. For the MLP model, the sequential model API was chosen, while dense was chosen for the layers. The initial weight for the model was set with HeUniform. Different combinations of hyperparameters, such as the number of hidden layers, number of nodes, and activation function, were tried for the model. The performance of the model stopped improving significantly after the addition of the fifth hidden layer. As for the number of nodes, there were 26 nodes in the input layer. The number of nodes for each hidden layer was 200, 150, 100, 50, and 9 for the last layer. Softmax was chosen as the activation function for the output layer, while the performance of ReLU and tanh were compared for the hidden layers of the model. The former showed an overall better result. The model was then compiled; the loss function was set to categorical hinge loss, Adam was chosen as the optimizer, and the learning rate was set to 0.001. The model was built using the Keras Classifier class. For the training process, LOOCV was conducted to estimate the performance of the model. Because the dataset used in this study was small, LOOCV was chosen to avoid overfitting (Pasini, 2015). This way, we could assess whether the model could generalize patterns in the training set despite the limited amount of data points. LOOCV was done using the function from the scikit-learn library. The epoch and batch size were set to 40 and 32, respectively.

The average score from LOOCV was used to evaluate the performance of the model. The prediction results from LOOCV were compared with the actual data. The accuracy was computed. A confusion matrix was created to see the classification performance of the model. The visualization was made using the Matplotlib library. The metrics used in this study were precision, recall, and F1 score. These metrics were calculated from rates of true positive, true negative, false positive, and false negative (Equations 1-3).

$$Precision = \frac{true\ positives}{true\ positives + false\ positives} \tag{1}$$

$$Recall = \frac{true\ positives}{true\ positives + false\ negatives} \tag{2}$$

$$F1\ score = 2 \times \frac{precision\ \times recall}{precision + recall} \tag{3}$$

The F1 score was computed for each class. From this score, the macro average and weighted average F1 score can be calculated.

RESULTS AND DISCUSSION

This study utilized the ANN method for generating a chatbot system using the symptom and comorbidities as the input and the medical specialist as the output. In total, 129 data of common symptoms, comorbidities, and the corresponding medical specialist were collected. After removing missing and ambiguous values, the final dataset consisted of 111 instances. In order to handle the missing and ambiguous values, we checked and manually filtered the data by eye. It is time-consuming, but having a good data portion in the training session will likely positively affect the testing results.

For the inputs, there were 18 symptoms and 8 comorbidities included in the dataset: chest discomfort, fainting, discomfort in the shoulder, neck, back, and jaw, shortness of breath, fractured bone, persistent headache, cuts, foreign material stuck, skin infection, the problem with vision, fever, diarrhea, cold sweat, back pain, weakness, trouble in speaking seeing and walking, vomit, and stomachache for common symptoms; hypertension, diabetes, heart disease, liver disease, stroke, tumor/cancer, respiration disease, and gastric for common comorbidities.

Meanwhile, there are nine types of medical specialties in the dataset. These medical specialties are internists, neurologists, cardiologists, ophthalmologists, pulmonologists, orthopedists, general surgeons, dermatologists, and Ear, Nose, and Throat/ENT specialists.

The percentage of data for each medical specialty can be seen in Table 1. Compared to other medical specialties, internists had the highest amount of data, while ENT had the lowest. Three medical specialties—internist, neurologist, and cardiologist—made up more than 70% of the dataset. On the other hand, each of the rest of the medical specialties made up less than 10%. Some solutions are suggested for the class imbalance problem in multi-class classification. These solutions include data resampling (oversampling or undersampling) and changing learning algorithms (Agrawal et al., 2015; Koziarski et al., 2020; So & Valdez, 2021; Tanha et al., 2020). Another important suggestion is choosing the most suitable performance evaluation method (Alejo et al., 2013; Luque et al., 2019).

Table 1
Percentage of data for each medical specialty

Medical Specialty	Percentage of Data (%)
SpPD / Internist	28.8
SpS / Neurologist	25.2
SpJP / Cardiologist	21.6
SpM / Ophthalmologist	8.1
Paru / Pulmonologist	5.4
SpOT / Orthopedist	4.5
SpKK / Dermatologist	2.7
SpB / General Surgeon	2.7
THT / ENT	0.9

The architecture of the best model in this study can be seen in Figure 1. It is an MLP model with five hidden layers, with the Rectified linear unit (ReLU) as the activation function for these layers. ReLU is considered the most popular activation function in deep learning (Cao et al., 2018). On the output layer, the chosen activation

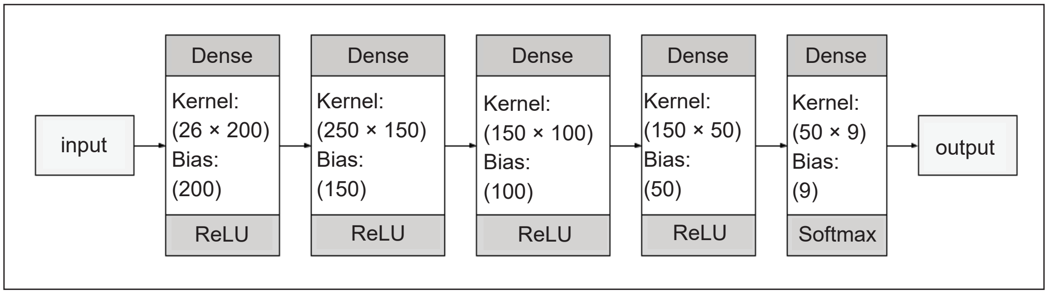

Figure 1. Architecture of the model

function was softmax. This activation function was also chosen because only one correct (medical specialty) answer is needed. After the model was compiled, the predictions made by the model during the LOOCV iterations with the actual data were compared. The result showed that the model has about 85.6% accuracy.

Because of the class imbalance, accuracy might not be the best metric to evaluate the model performance. Accuracy is the result of dividing the number of correct predictions by the total number of predictions without considering the number of correct classifications for each class. Accuracy with an imbalanced dataset can be misleading, as the score would not show how well the model performs at classifying the underrepresented classes. Kulkarni et al. (2021) suggested that metrics based on a confusion matrix can be more useful for evaluating the performance of models trained with an imbalanced dataset. These metrics include precision, recall, and F1 score.

Although the accuracy score was good, the model only seemed to classify some medical specialties correctly. It can be observed in the visualization of the confusion matrix (Figure 2). The model can classify data of internists, neurologists, and cardiologists really well; it can classify all of the cardiologist data correctly. However, the model showed poorer performance for classes with fewer data. The model seemed to misclassify all data for the two most underrepresented classes, ENT and surgeon. The macro average and weighted average F1 scores were computed to evaluate the performance of the model. Due to the class imbalance, the weighted average F1 score was higher than the macro average. The values for these metrics were about 0.59 for the macro average and about 0.84 for the weighted average.

The precision, recall, and F1 score were obtained from the confusion matrix. The scores for each class can be seen in Table 2. High precision, recall, and F1 scores were shown for the three classes with the highest amount of data. Neurologist, the class with the most data, achieved an F1 score of 0.91. Another class, internist, also achieved the same score. A perfect score on all three metrics was observed for the cardiologist class. The model also seemed able to classify all orthopedist data correctly, but classify some data of other classes as an orthopedist, resulting in perfect recall but a lower precision score. The pulmonologist class can see the opposite, as the precision is higher than recall. Most data classified as pulmonologists were correct, but the model misclassified 40% of pulmonologist data. For ophthalmologist and dermatologist classes, the model showed poor performance, with an F1 score of 0.53 and 0.40, respectively.

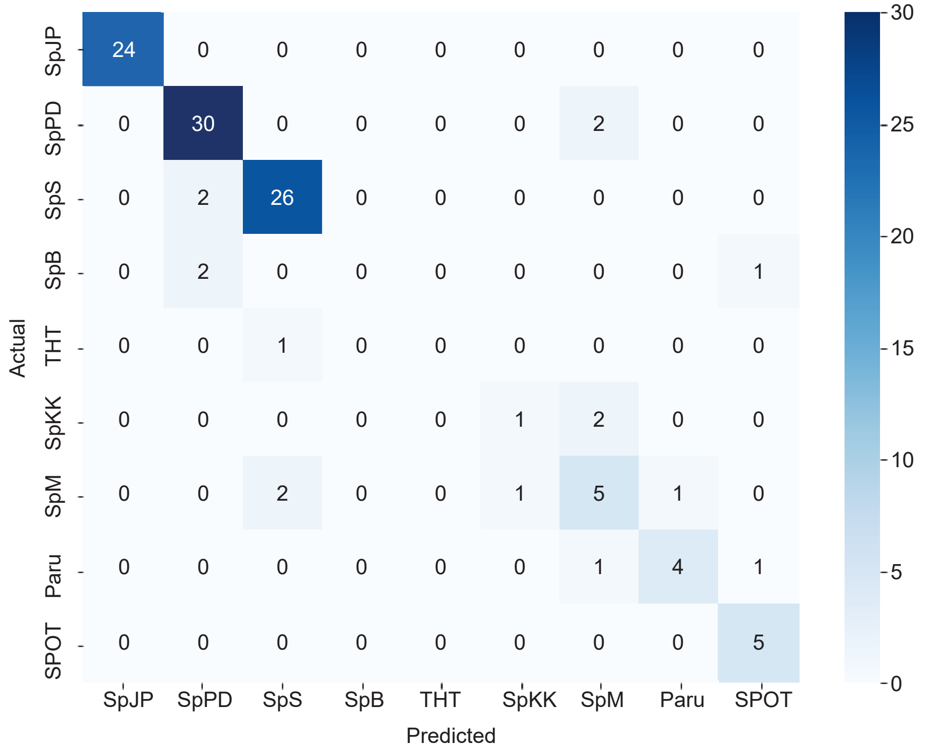

Figure 2. Heat map of the confusion matrix

Table 2
Precision, recall, and F1 score for each medical specialty

Medical Specialty	Precision	Recall	F1 score
SpPD	0.88	0.94	0.91
SpS	0.90	0.93	0.91
SpJP	1.00	1.00	1.00
SpM	0.50	0.56	0.53
Paru	0.80	0.60	0.73
SpOT	0.71	1.00	0.83
SpKK	0.50	0.33	0.40
SpB	0.00	0.00	0.00
THT	0.00	0.00	0.00

The class imbalance in the dataset affected the performance of the model. Lee et al. (2021) also used an imbalanced dataset in their study in which they developed a deep learning-based NLP for medical specialty recommendations. Two models in their study achieved a macro average F1 score of 0.74 and 0.77. Compared to this study, the dataset they used had more instances; their dataset contained 51,134 data points. The better classification performance despite the class imbalance can be due to the dataset size and the use of a pre-trained model.

Besides the small and imbalanced dataset, this study also has other limitations. Firstly, it would be difficult for the model to recommend patients with more complicated conditions: who have several symptoms and at the same time also have multiple comorbidities. The next limitation is that we did not consider the sub-specialist. For future works, we will consider using more advanced AI techniques, such as convolutional neural networks (CNN).

CONCLUSION

In this study, an ANN model was developed for medical specialty recommendations. ANN can be a useful method for classification, even for smaller datasets. The model proposed in this study can correctly classify classes of medical specialties with 24 to 32 data points, which shows that ANN can be suitable for small datasets. However, the class imbalance in the dataset caused the model to perform well at predicting only 5 out of 9 medical specialties. More data is needed, especially for classes such as ENT, Surgeon, Ophthalmologist, and Dermatologist.

ACKNOWLEDGMENT

The project is funded under Hibah Keilmuan Dirjen Dikti 2022, Indonesia. The authors thank the Department of Research and Community Service of Indonesia International Institute for Life Sciences for supporting this research.

REFERENCES

Agrawal, A., Viktor, H. L., & Paquet, E. (2015, November 12-14). SCUT: Multi-class imbalanced data classification using SMOTE and cluster-based undersampling. [Paper presentation]. *International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management* (IC3K), Lisbon, Portugal. https://doi.org/10.5220/0005595502260234

Aguiar, F. S., Torres, R. C., Pinto, J. V. F., Kritski, A. L., Seixas, J. M., & Mello, F. C. Q. (2016). Development of two artificial neural network models to support the diagnosis of pulmonary tuberculosis in hospitalized patients in Rio de Janeiro, Brazil. *Medical & Biological Engineering & Computing*, *54*(11), 1751-1759. https://doi.org/10.1007/s11517-016-1465-1

Alejo, R., Antonio, J. A., Valdovinos, R. M., & Pacheco-Sánchez, J. H. (2013). Assessments Metrics for Multi-class Imbalance Learning: A Preliminary Study. In J. A. Carrasco-Ochoa, J. F. Martinex-Trinidad, J. S.

Rodriuez & G. S. D. Baja (Eds.), *Pattern Recognition: 5th Mexican Conference, MCPR 2013, Querétaro, Mexico Proceedings 5* (pp. 335-343). Springer. https://doi.org/10.1007/978-3-642-38989-4_34

Amato, F., López, A., Peña-Méndez, E. M., Vaňhara, P., Hampl, A., & Havel, J. (2013). Artificial neural networks in medical diagnosis. *Journal of Applied Biomedicine*, *11*(2), 47-58. https://doi.org/10.2478/v10136-012-0031-x

Borghi, P. H., Zakordonets, O., & Teixeira, J. P. (2021). A COVID-19 time series forecasting model based on MLP ANN. *Procedia Computer Science*, *181*, 940-947. https://doi.org/10.1016/j.procs.2021.01.250

Buscema, P. M., Gitto, L., Russo, S., Marcellusi, A., Fiori, F., Maurelli, G., Massini, G., & Mennini, F. S. (2017). The perception of corruption in health: AutoCM methods for an international comparison. *Quality & Quantity*, *51*(1), 459-477. https://doi.org/10.1007/s11135-016-0315-4

Cao, C., Liu, F., Tan, H., Song, D., Shu, W., Li, W., Zhou, Y., Bo, X., & Xie, Z. (2018). Deep learning and its applications in biomedicine. *Genomics, Proteomics & Bioinformatics*, *16*(1), 17-32. https://doi.org/10.1016/j.gpb.2017.07.003

Casagranda, I., Costantino, G., Falavigna, G., Furlan, R., & Ippoliti, R. (2016). Artificial neural networks and risk stratification models in emergency departments: The policy maker's perspective. *Health Policy*, *120*(1), 111-119. https://doi.org/10.1016/j.healthpol.2015.12.003

Chai, S. S., Cheah, W. L., Goh, K. L., Chang, Y. H. R., Sim, K. Y., & Chin, K. O. (2021). A multilayer perceptron neural network model to classify hypertension in adolescents using anthropometric measurements: A cross-sectional study in Sarawak, Malaysia. *Computational and Mathematical Methods in Medicine*, *2021*, Article 2794888. https://doi.org/10.1155/2021/2794888

da Silva, I. N., Spatti, D. H., Flauzino, R. A., Liboni, L. H. B., & Alves, S. F. D. R. (2017). Artificial neural network architectures and training processes. In *Artificial Neural Networks: A Practical Course* (pp. 21-28). Springer International Publishing. https://doi.org/10.1007/978-3-319-43162-8_2

Feng, S., Zhou, H., & Dong, H. (2019). Using deep neural network with small dataset to predict material defects. *Materials & Design*, *162*, 300-310. https://doi.org/10.1016/j.matdes.2018.11.060

Ippoliti, R., Falavigna, G., Zanelli, C., Bellini, R., & Numico, G. (2021). Neural networks and hospital length of stay: An application to support healthcare management with national benchmarks and thresholds. *Cost Effectiveness and Resource Allocation*, *19*(1), Article 67. https://doi.org/10.1186/s12962-021-00322-3

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, *2*(4), 230-243. https://doi.org/10.1136/svn-2017-000101

Koziarski, M., Woźniak, M., & Krawczyk, B. (2020). Combined cleaning and resampling algorithm for multi-class imbalanced data with label noise. *Knowledge-Based Systems, 204*, Article 106223. https://doi.org/10.1016/j.knosys.2020.106223

Kulkarni, A., Chong, D., & Batarseh, F. A. (2021). Foundations of data imbalance and solutions for a data democracy. In F. A. Batarseh & R. Yang (Eds.), *Data Democracy* (pp. 83-106). Academic Press. https://doi.org/10.1016/B978-0-12-818366-3.00005-8

Kumar, A., Prakash, U. M., & Sharma, G. K. (2021). Disease prediction and doctor recommendation system using machine learning approaches. *International Journal for Research in Applied Science and Engineering Technology*, *9*(VII), 34-44. https://doi.org/10.22214/ijraset.2021.36234

Lee, H., Kang, J., & Yeo, J. (2021). Medical specialty recommendations by an artificial intelligence chatbot on a smartphone: Development and deployment. *Journal of Medical Internet Research*, *23*(5), Article e27460. https://doi.org/10.2196/27460

Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, *91*, 216-231. https://doi.org/10.1016/j.patcog.2019.02.023

Olson, M., Wyner, A., & Berk, R. (2018, December 2-8). *Modern neural networks generalize on small data sets*. [Paper presentation]. Conference on Neural Information Processing Systems (NeurIPS), Montreal, Canada.

Pasini, A. (2015). Artificial neural networks for small dataset analysis. *Journal of Thoracic Disease*, *7*(5), 953-960. https://doi.org/10.3978/j.issn.2072-1439.2015.04.61

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. In L. Liu & M. T. Ozsu (Eds.), *Encyclopedia of Database Systems* (pp. 532-538). Springer. https://doi.org/10.1007/978-0-387-39940-9_565

Rémy, N. M., Martial, T. T., & Clémentin, T. D. (2018). The prediction of good physicians for prospective diagnosis using data mining. *Informatics in Medicine Unlocked*, *12*, 120-127. https://doi.org/10.1016/j.imu.2018.07.005

Shahid, N., Rappon, T., & Berta, W. (2019). Applications of artificial neural networks in health care organizational decision-making: A scoping review. *PloS One*, *14*(2), Article e0212356. https://doi.org/10.1371/journal.pone.0212356

Silitonga, P., Bustamam, A., Muradi, H., Mangunwardoyo, W., & Dewi, B. E. (2021). Comparison of dengue predictive models developed using artificial neural network and discriminant analysis with small dataset. *Applied Sciences*, *11*(3), Article 943. https://doi.org/10.3390/app11030943

So, B., & Valdez, E. A. (2021). *The SAMME.C2 algorithm for severely imbalanced multi-class classification*. ArXiv. https://doi.org/10.48550/arXiv.2112.14868

Tanha, J., Abdi, Y., Samadi, N., Razzaghi, N., & Asadpour, M. (2020). Boosting methods for multi-class imbalanced data classification: an experimental review. *Journal of Big Data*, *7*(1), Article 70. https://doi.org/10.1186/s40537-020-00349-y

Webb, G. I., Sammut, C., Perlich, C., Horváth, T., Wrobel, S., Korb, K. B., Noble, W. S., Leslie, C., Lagoudakis, M. G., Quadrianto, N., Buntine, W. L., Quadrianto, N., Buntine, W. L., Getoor, L., Namata, G., Getoor, L., Han, X. J. J., Ting, J. A., Vijayakumar, S., … & Raedt, L. D. (2011). Leave-One-Out Cross-Validation. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning* (pp. 600-601). Springer. https://doi.org/10.1007/978-0-387-30164-8_469

Yao, L., Zhong, Y., Wu, J., Zhang, G., Chen, L., Guan, P., Huang, D., & Liu, L. (2019). Multivariable logistic regression and back propagation artificial neural network to predict diabetic retinopathy. *Diabetes, Metabolic Syndrome and Obesity*, *12*, 1943-1951. https://doi.org/10.2147/DMSO.S219842