

Research Article

Prediction Model between Serum Vitamin D and Neurological Deficit in Cerebral Infarction Patients Based on Machine Learning

Hailiu Zhang , Guotao Yang , and Aiqin Dong 

Department of Neurology, Cangzhou Central Hospital, Cangzhou 061000, China

Correspondence should be addressed to Aiqin Dong; aiqin_drdong@163.com

Received 19 March 2022; Revised 27 May 2022; Accepted 15 June 2022; Published 28 June 2022

Academic Editor: Xue Fei Deng

Copyright © 2022 Hailiu Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. Vitamin D is associated with neurological deficits in patients with cerebral infarction. This study uses machine learning to evaluate the prediction model's efficacy of the correlation between vitamin D and neurological deficit in patients with cerebral infarction. **Methods.** A total of 200 patients with cerebral infarction admitted to the Department of Neurology of our hospital from July 2018 to June 2019 were selected. The patients were randomly divided into a training set ($n = 140$) and a test set ($n = 60$) in a 7 : 3 ratio. The prediction model is constructed from the training set's data, and the model's prediction effect was evaluated by test set data. The area under the receiver operator characteristic curve was used to assess the prediction efficiency of models. **Results.** In the training set, the area under the curve (AUC) of the logistic regression model and XGBoost algorithm model was 0.727 (95% CI: 0.601~0.854) and 0.818 (95% CI: 0.734~0.934), respectively. While in the test set, the AUC of the logistic regression model and XGBoost algorithm model was 0.761 (95% CI: 0.640~0.882) and 0.786 (95% CI: 0.670~0.902), respectively. **Conclusion.** The prediction model of the correlation between vitamin D and neurological deficit in patients with cerebral infarction based on machine learning has a good prediction efficiency.

1. Introduction

With the improvement of social and economic living standards and the aging of the population, the incidence of acute cerebral infarction is rising [1, 2]. Cerebral infarction has a high disability rate, high mortality rate, and high recurrence rate, which seriously harms the health of the social population and brings a heavy economic burden and medical burden to the family and society [3, 4]. Therefore, the treatment and prevention of stroke to reduce its incidence, mortality, disability rate, and cost of medical treatment has become a fundamental goal of the medical circle.

Vitamin D is a fat-soluble steroid derivative [5]. Vitamin D is hydroxylated in the body to form steroid hormones that regulate metabolism through genomic and nongenomic pathways [6, 7]. In addition to the traditional role of osteoporosis prevention and calcium metabolism

[8], a growing number of studies have linked vitamin D deficiency with cardiovascular disease [9], kidney disease, and infections [10]. Vitamin D metabolism and brain diseases are also hot topics. A large number of studies have suggested that low vitamin D levels increase both the incidence of stroke and the incidence of adverse outcomes in stroke patients [11].

The 25-hydroxyvitamin D₃ (25(OH)D₃) is an activated vitamin D and is the best indicator of vitamin D levels in the body [12]. In recent years, it has been reported that serum 25(OH)D₃ level is associated with neurological recovery of acute cerebral infarction, and patients with normal 25(OH)D₃ levels have a better prognosis than those with 25(OH)D₃ deficiency [13, 14]. With the development of biomedicine, the application of machine learning in the medical field is increasing gradually [15–17]. Therefore, this study constructed infarction based on machine learning, providing

a method for clinical diagnosis and treatment of neurological deficit in patients with cerebral infarction.

2. Methods

2.1. Research Object. A total of 200 patients with cerebral infarction admitted to the Department of Neurology of our hospital from July 2018 to June 2019 were selected. Inclusion criteria were as follows: (1) clinical diagnosis of acute cerebral infarction: meeting the diagnostic criteria set by the 4th National Academic Conference on Cerebrovascular Diseases and (2) the onset of all patients was within seven days. Exclusion criteria were as follows: (1) patients with severe heart, liver, and kidney diseases; (2) patients with previous neurological diseases and severe sequelae; and (3) patients who took drugs affecting vitamin D metabolism within one week.

2.2. Research Methods. Demographic information, past history, personal and family history, and whether patients had taken vitamin D in the last 1 week were collected. All patients underwent blood pressure measurement, nervous system physical examination, and brain imaging examination after admission. Fasting venous blood was taken on the second day of admission to determine the serum 25(OH)D₃, triglyceride (TG), total cholesterol (TC), low-density lipoprotein (LDL-C), high-density lipoprotein (HDL-C), and other related laboratory indicators. In addition, all subjects were assessed on the National Institute of Health Stroke Scale (NIHSS) on the day of admission. The scale included consciousness, sensation, visual field, gaze, facial paralysis, upper and lower limb movement, ataxia movement, language, and dysarthria. According to their actual situation and physical examination results, the total score of each item is evaluated one by one, which is the NIHSS score of the research object. All the above were carried out under the supervision of physicians with 5 years of clinical experience. Scoring principle is as follows: (1) only the first reaction of the research object is recorded for each item; (2) record the actual ability of the subjects, rather than doctors' subjective opinion; (3) the patients were examined and recorded at the same time without giving any hints to the subjects; and (4) if some items cannot be evaluated, record the highest score of the item. According to NIHSS scoring criteria, NIHSS < 7 scores are categorized as mild injury, NIHSS 7-14 scores are categorized as moderate injury, and NIHSS > 14 scores are categorized as severe injury.

2.3. Machine Learning Model Construction. A data set of 200 patients was randomly divided into a training set ($n = 140$) and a test set ($n = 60$) in a 7:3 ratio. Firstly, the training set data is preprocessed to normalize the data. Then, the prediction model is constructed from the data of the training set, and the model is trained. The prediction effect of the model was evaluated by test set data. Then, the predictive efficiency of the model is evaluated. The specific flow chart is shown in Figure 1. Patients in the training set were divided

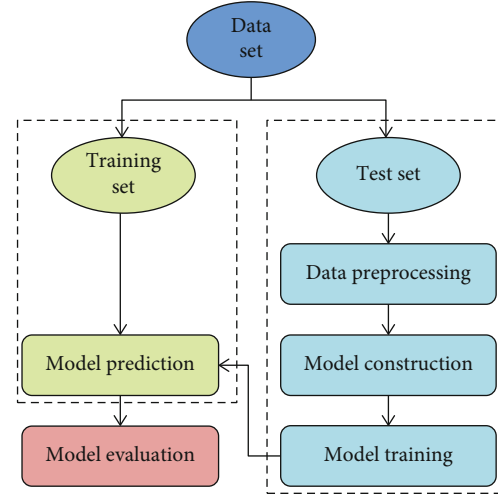


FIGURE 1: Flow chart of machine learning model construction.

into mild group ($n = 99$), moderate group ($n = 31$), and severe group ($n = 10$) according to the score of the NIHSS.

2.4. Construction of Logistic Regression Model. Logistic regression, also known as logistic regression analysis, is a generalized linear predictive regression model often used in data mining, disease diagnosis, and economic forecasting. Logistic regression is the most popular binary data model, and the general logistic regression model has the following forms:

$$\text{Log} \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \alpha + \beta x. \quad (1)$$

The 1-dimensional logistic model has the following form:

$$\begin{aligned} P(Y = y) &= \mu^y (1 - \mu)^{1-y} = (1 - \mu) \exp \left(y \log \frac{\mu}{1 - \mu} \right) \\ &= \exp \left\{ y\theta - \log \left(1 + e^\theta \right) \right\}, \\ y &= 0, 1, \theta = \log \frac{\mu}{1 - \mu}. \end{aligned} \quad (2)$$

The natural connection function is $\theta = z^T \beta$, $\mu = e^{z^T \beta} / (1 + e^{z^T \beta})$.

2.5. Construction of XGBoost Model. The base learner of the XGBoost algorithm is mainly the classification and regression trees (CART), which can effectively improve the overfitting problem of the single tree model [18]. The main idea is to select some sample features to generate a base learner and continuously fit the previous residuals to minimize the objective function. XGBoost algorithm can be regarded as an additional model composed of K trees [19]. The tree

model used in this paper is a regression tree, and the specific formula is as follows:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F, \quad (3)$$

where \hat{y}_i is the prediction of sample x_i , i indicates the serial number of the input sample, K is the synthesis of threes, and F is the set space of all regression trees.

The modeling process of XGBoost keeps the original model unchanged and takes the error generated by the last prediction as a reference to build the next tree.

Initialization:

$$\hat{y}_i^{(0)} = 0. \quad (4)$$

Add the first tree to the model:

$$\hat{y}_i^{(1)} = f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i). \quad (5)$$

Add the n th first tree to the model:

$$\hat{y}_i^{(n)} = \sum_{k=1}^n f_k(x_i) = \hat{y}_i^{(n-1)} + f_n(x_i), \quad (6)$$

where $\hat{y}_i^{(n)}$ is the predicted value of the i th sample at the n th time, which retains the model prediction result of $n-1$ time and adds a new function $f_n(x_i)$. New functions are added in each round to minimize the loss function. At this point, the loss function is

$$L(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)). \quad (7)$$

First, the clinical data and laboratory results of patients were predicted by logistic regression. Then, the XGBoost model was tuned by adjusting the weight of leaf nodes and the depth of tree model. The XGBoost parameters were adjusted by the fivefold crossover method to obtain the best prediction model, and finally, the feature selection was carried out. The specific process is shown in Figure 2.

2.6. Statistical Analysis. SPSS 20.0 software was used for statistical analysis. The measurement data were tested for Shapiro-Wilk normality. The normal distribution was shown by the mean \pm standard deviation table, and the nonnormal distribution was expressed by median (interquartile spacing). The Levene method was used to test homogeneity of variance. Bivariate analysis of measurement data: two independent sample t -tests were used when normal variance was uniform, t -test when normal variance is not uniform, and Wilcoxon rank-sum test for nonnormal distribution. Significance test levels were all $P < 0.05$ which were statistically different.

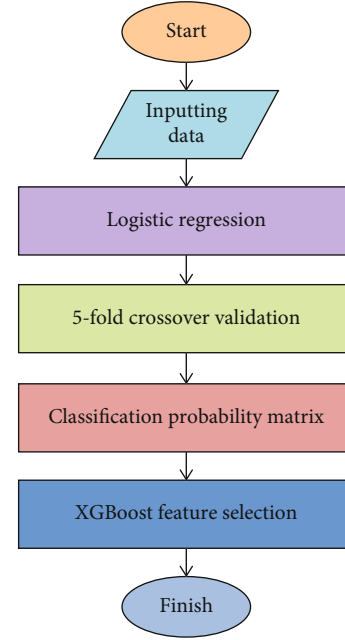


FIGURE 2: Flow chart based on logistic regression and XGBoost prediction model.

TABLE 1: Comparison of clinical data and laboratory results of patients in training set and test set [n (%)].

Indicator	Training set ($n = 140$)	Test set ($n = 60$)	χ^2	P
Male	51 (36.4%)	22 (36.7%)	0.001	0.974
Smoking	58 (41.4%)	24 (40.0%)	0.035	0.851
Hypertension	95 (67.9%)	40 (66.7%)	0.027	0.869
Diabetes	45 (32.1%)	19 (31.7%)	0.004	0.947

3. Results

3.1. Comparison of Clinical Data and Laboratory Results of Patients in Different Sets. Clinical data and laboratory results of patients in the training and test sets were compared. The results showed that there was no significant difference in gender, age, hypertension, diabetes, smoking history, and other laboratory indicators, such as TC, TG, LDL-C, HDL-C, and 25(OH)D₃ between the two groups ($P > 0.05$). The specific results are shown in Table 1 and Figure 3.

3.2. Comparison of 25(OH)D₃ Levels under Different Neurological Deficit in Training Set. Patients in the training set were divided into mild injury group ($n = 99$), moderate injury group ($n = 31$), and severe injury group ($n = 10$) according to the score of the NIHSS. The serum 25(OH)D₃ levels in the three groups decreased gradually from the mild group to the severe group, and the differences were statistically significant ($P < 0.05$) (Figure 4).

3.3. Evaluation of the Effectiveness of Prediction Models. The area under the receiver operator characteristic (ROC) curve of the logistic regression model in the training set was

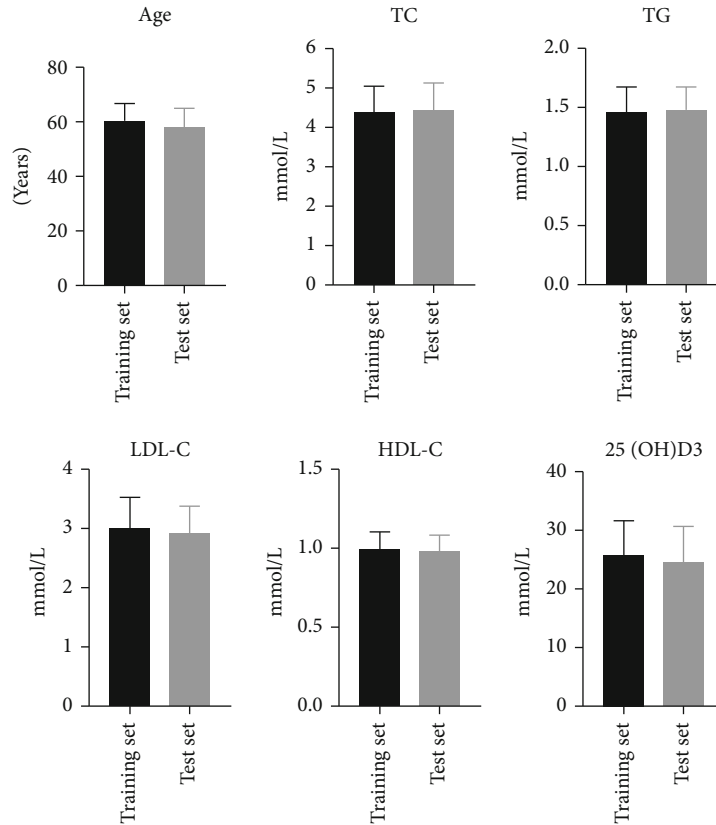


FIGURE 3: Comparison of clinical data and laboratory results of patients in training set and test set. There was no statistically significant difference between the two sets ($P > 0.05$).

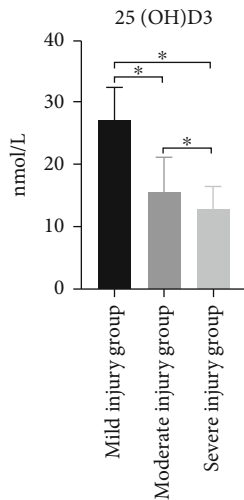


FIGURE 4: Comparison of 25(OH)D₃ levels in different groups with neurological deficit. *: $P < 0.05$.

0.727 (95% CI: 0.601~0.854). The area under the curve (AUC) of the XGBoost algorithm model is 0.818 (95% CI: 0.734~0.934). In addition, the area under the curve of the logistic regression model in the test set was 0.761 (95% CI: 0.640~0.882). The AUC of the XGBoost algorithm model is 0.786 (95% CI: 0.670~0.902). The detailed ROC curve was shown in Figure 5.

4. Discussion

Cerebral infarction refers to the clinical syndrome of cerebral blood supply disorders caused by various reasons, resulting in local cerebral tissue ischemia, hypoxic necrosis, and corresponding neurological defects [20]. Cerebral infarction is one of the three major diseases threatening human health and survival [21]. Cerebral infarction has a high incidence, prevalence, recurrence rate, disability rate, and mortality rate, which brings a great burden to families and society. The majority of cerebral infarction is on the rise in China, which has seriously harmed people's health [22]. Therefore, the treatment and prevention of stroke to reduce its incidence, mortality, disability rate, and lower cost of medical treatment has become a significant goal of the medical circle.

The 25(OH)D₃ has an extensive biological effect. Besides classic calcium-phosphorus metabolism regulation, anti-inflammatory, immune regulation, lipid metabolism, and cell differentiation, it also inhibits the renin-angiotensin system and affects blood pressure and insulin secretion. This effect can protect target organs such as the heart, blood vessels, lower blood pressure, and blood sugar and slow the hardening of the arteries [23–25]. Recent studies at home and abroad suggest that a low 25(OH)D₃ level is an independent risk factor for cardiovascular and cerebrovascular diseases [26]. Patients with low 25(OH)D₃ levels were found to have a significantly increased risk of acute stroke, of which 25(OH)D₃ was a manageable risk factor [27]. Kiggundu et al. [28] collected 142

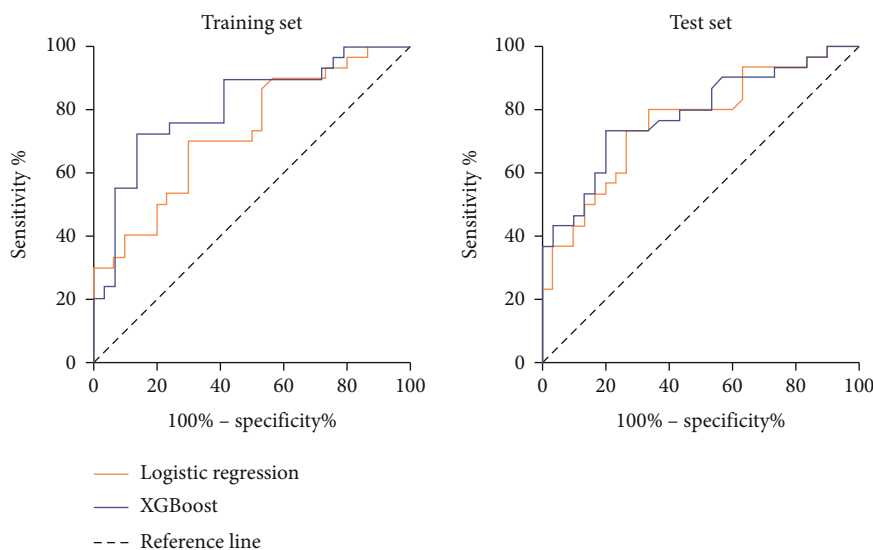


FIGURE 5: ROC curves of logistic regression model and XGBoost algorithm model in different sets.

patients with acute stroke from a central hospital in Kampala and detected a level of 25(OH)D₃ in their serum. The results showed that vitamin D deficiency was correlated with stroke. This is consistent with our study, and we found that serum 25 levels were associated with the level of neurological impairment in patients. The serum 25 levels were lowest in patients with severe neurological impairment.

The mechanism of the correlation between serum 25(OH)D₃ level and the condition and prognosis of patients with acute ischemic stroke is still unclear. Based on the previous studies, it can be summarized as follows: (1) serum 25(OH)D₃ has an anti-inflammatory effect, and 25(OH)D₃ can inhibit endoplasmic reticulum stress during inflammation and reduce the expression of monocyte chemoattractant protein 1. This reduces cholesterol deposition in macrophages and ultimately inhibits the formation of atherosclerotic plaques. At the same time, by binding VDR on immune cells to upregulate anti-inflammatory factors, downregulate the expression of inflammatory factors to play an anti-inflammatory effect [29]. (2) Animal experiments have confirmed that serum 25(OH)D₃ can improve hypercoagulability and antithrombogenesis by downregulating procoagulant tissue factor and upregulating the expression of thrombomodulin [30]. (3) In addition, animal experiments showed that 25(OH)D₃ had protective effects on focal cerebral ischemia-reperfusion injury, which may be related to reducing oxygen-free radical injury and promoting microvascular regeneration [31]. (4) Atif et al. [32] found that vitamin D can enhance the neuroprotective effect of P4 and reduce the volume of cerebral infarction in the middle cerebral artery infarction model. Vitamin D also regulates the synthesis of neurotransmitters such as acetylcholine, 5-hydroxytryptamine, and dopamine, affecting neurological function.

This study builds a prediction model based on logistic regression and the XGBoost algorithm, and the results show that the area under the curve of the logistic regression model in the test set was 0.761 (95% CI: 0.640~0.882). The area under the curve of the XGBoost algorithm model is 0.786

(95% CI: 0.670~0.902), which shows that the XGBoost model has better prediction performance than logistic regression and can achieve more accurate individual prediction. However, in the training set, there was no significant difference between the two prediction models. The possible reason is that although the XGBoost algorithm model has a unique advantage in dealing with nonlinear relations with high-dimensional variables, the effectiveness of the prediction model is also affected by the nature of variables, sample size, and other factors.

There are some limitations to this study. This study was a single-center cross-sectional retrospective study with a small sample. Further validation is needed in a multicenter prospective study with a large sample. In addition, the effectiveness of the prediction model is affected by the nature and number of variables, which could not play a role in dealing with nonlinear relations.

5. Conclusion

The prediction model of the correlation between vitamin D and neurological deficit in patients with cerebral infarction based on machine learning has good prediction efficiency, which can provide a clinical diagnosis and treatment method.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

The study was supported by the Hebei Medical Science Research Project Plan (Reference No: 20200308).

References

- [1] S. J. van Veluw, A. Y. Shih, E. E. Smith et al., "Detection, risk factors, and functional consequences of cerebral microinfarcts," *Lancet Neurology*, vol. 16, no. 9, pp. 730–740, 2017.
- [2] S. Liu, J. R. Wu, D. Zhang et al., "Comparative efficacy of chinese herbal injections for treating acute cerebral infarction: a network meta-analysis of randomized controlled trials," *BMC Complementary and Alternative Medicine*, vol. 18, no. 1, p. 120, 2018.
- [3] L. A. Perry, E. Berge, J. Bowditch et al., "Antithrombotic treatment after stroke due to intracerebral haemorrhage," *Cochrane Database of Systematic Reviews*, vol. 5, article Cd012144, 2017.
- [4] L. Jin, J. Zhou, W. Shi et al., "Effects of six types of aspirin combination medications for treatment of acute cerebral infarction in China: a network meta-analysis," *Journal of Clinical Pharmacy and Therapeutics*, vol. 44, no. 1, pp. 91–101, 2019.
- [5] S. W. Chang and H. C. Lee, "Vitamin D and health - the missing vitamin in humans," *Pediatrics and Neonatology*, vol. 60, no. 3, pp. 237–244, 2019.
- [6] M. F. Holick, "The vitamin D deficiency pandemic: approaches for diagnosis, treatment and prevention," *Reviews in Endocrine & Metabolic Disorders*, vol. 18, no. 2, pp. 153–165, 2017.
- [7] A. R. Martineau, D. A. Jolliffe, L. Greenberg et al., "Vitamin D supplementation to prevent acute respiratory infections: individual participant data meta-analysis," *Health Technology Assessment*, vol. 23, no. 2, pp. 1–44, 2019.
- [8] C. Liu, X. Kuang, K. Li, X. Guo, Q. Deng, and D. Li, "Effects of combined calcium and vitamin D supplementation on osteoporosis in postmenopausal women: a systematic review and meta-analysis of randomized controlled trials," *Food & Function*, vol. 11, no. 12, pp. 10817–10827, 2020.
- [9] J. E. Manson, N. R. Cook, I. M. Lee et al., "Vitamin D supplements and prevention of cancer and cardiovascular disease," *The New England Journal of Medicine*, vol. 380, no. 1, pp. 33–44, 2019.
- [10] G. Jean, J. C. Souberbielle, and C. Chazot, "Vitamin D in chronic kidney disease and dialysis patients," *Nutrients*, vol. 9, no. 4, p. 328, 2017.
- [11] R. Zhou, M. Wang, H. Huang, W. Li, Y. Hu, and T. Wu, "Lower vitamin D status is associated with an increased risk of ischemic stroke: a systematic review and meta-analysis," *Nutrients*, vol. 10, no. 3, p. 277, 2018.
- [12] M. Sosa Henríquez and M. J. G. de Tejada Romero, "Cholecalciferol or calcifediol in the management of vitamin D deficiency," *Nutrients*, vol. 12, no. 6, p. 1617, 2020.
- [13] A. El-Sharkawy and A. Malki, "Vitamin D signaling in inflammation and cancer: molecular mechanisms and therapeutic implications," *Molecules*, vol. 25, no. 14, p. 3219, 2020.
- [14] J. Guo, J. A. Lovegrove, and D. I. Givens, "25(OH)D₃-enriched or fortified foods are more efficient at tackling inadequate vitamin D status than vitamin D₃," *The Proceedings of the Nutrition Society*, vol. 77, no. 3, pp. 282–291, 2018.
- [15] H. Wang, B. Song, N. Ye et al., "Machine learning-based multi-parametric mri radiomics for predicting the aggressiveness of papillary thyroid carcinoma," *European Journal of Radiology*, vol. 122, article 108755, 2020.
- [16] S. Deb, Z. Tian, S. Fong, R. Wong, R. Millham, and K. K. L. Wong, "Elephant search algorithm applied to data clustering," *Soft Computing*, vol. 22, no. 18, pp. 6035–6046, 2018.
- [17] K. Wong, G. Fortino, and D. Abbott, "Deep learning-based cardiovascular image diagnosis: a promising challenge," *Future Generation Computer Systems*, vol. 110, pp. 802–811, 2020.
- [18] K. Saqib, A. F. Khan, and Z. A. Butt, "Machine learning methods for predicting postpartum depression: scoping review," *JMIR Mental Health*, vol. 8, no. 11, article e29838, 2021.
- [19] N. Wang, M. Zeng, Y. Li, F. X. Wu, and M. Li, "Essential protein prediction based on node2vec and xgboost," *Journal of Computational Biology*, vol. 28, no. 7, pp. 687–700, 2021.
- [20] H. Suzuki, H. Kanamaru, F. Kawakita, R. Asada, M. Fujimoto, and M. Shiba, "Cerebrovascular pathophysiology of delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage," *Histology and Histopathology*, vol. 36, no. 2, pp. 143–158, 2021.
- [21] D. Mozaffarian, E. J. Benjamin, A. S. Go, D. K. Arnett, M. J. Blaha, and M. Cushman, "Executive summary: heart disease and stroke statistics2016 update: a report from the american heart association," *Circulation*, vol. 133, no. 4, pp. 447–454, 2016.
- [22] Y. X. Shang, L. F. Yan, E. M. Cornett, A. D. Kaye, G. B. Cui, and H. Y. Nan, "Incidence of cerebral infarction in Northwest China from 2009 to 2018," *Cureus*, vol. 13, no. 8, article e17576, 2021.
- [23] S. Pilz, A. Zittermann, R. Obeid et al., "The role of vitamin D in fertility and during pregnancy and lactation: a review of clinical data," *International Journal of Environmental Research and Public Health*, vol. 15, no. 10, p. 2241, 2018.
- [24] I. R. Reid and M. J. Bolland, "Controversies in medicine: the role of calcium and vitamin D supplements in adults," *The Medical Journal of Australia*, vol. 211, no. 10, pp. 468–473, 2019.
- [25] V. Rai and D. K. Agrawal, "Role of vitamin D in cardiovascular diseases," *Endocrinology and Metabolism Clinics of North America*, vol. 46, no. 4, pp. 1039–1059, 2017.
- [26] Emerging Risk Factors Collaboration, "Estimating dose-response relationships for vitamin D with coronary heart disease, stroke, and all-cause mortality: observational and mendelian randomisation analyses," *The Lancet Diabetes and Endocrinology*, vol. 9, no. 12, pp. 837–846, 2021.
- [27] G. Muscogiuri, L. Barrea, B. Altieri et al., "Calcium and vitamin D supplementation. Myths and realities with regard to cardiovascular risk," *Current Vascular Pharmacology*, vol. 17, no. 6, pp. 610–617, 2019.
- [28] D. S. Kiggundu, E. Mutebi, D. Kibirige et al., "Vitamin D deficiency and its characteristics among patients with acute stroke at a national referral hospital in Kampala Uganda," *BMC Endocrine Disorders*, vol. 15, no. 1, pp. 1–8, 2015.
- [29] A. E. Riek, J. Oh, I. Darwech et al., "Vitamin D₃ supplementation decreases a unique circulating monocyte cholesterol pool in patients with type 2 diabetes," *The Journal of Steroid Biochemistry and Molecular Biology*, vol. 177, pp. 187–192, 2018.
- [30] K. Zhou, B. Enkhjargal, Z. Xie et al., "Dihydrolipoic acid inhibits lysosomal rupture and nlrp3 through lysosome-associated membrane protein-1/calcium/calmodulin-dependent protein kinase ii/tak1 pathways after subarachnoid hemorrhage in rat," *Stroke*, vol. 49, no. 1, pp. 175–183, 2018.
- [31] K. Zhou, B. Enkhjargal, J. Mo et al., "Dihydrolipoic acid enhances autophagy and alleviates neurological deficits after subarachnoid hemorrhage in rats," *Experimental Neurology*, vol. 342, article 113752, 2021.
- [32] F. Atif, S. Yousuf, C. Espinosa-Garcia, W. A. C. Harris, and D. G. Stein, "Post-ischemic stroke systemic inflammation: immunomodulation by progesterone and vitamin D hormone," *Neuropharmacology*, vol. 181, article 108327, 2020.