Scholar Publishing Group
Machine Learning Theory and Practice
https://doi.org/10.38007/ML.2022.030106
ISSN 2790-0983 Vol. 3, Issue 1: 44-53

Dynamic Assessment of Agricultural Drought Risk Based on Machine Learning

Kothapa Lakshmi*

Indira Gandhi Delhi Technical University for Women, India

**corresponding author*

Keywords: Machine Learning, Random Forest Algorithm, Agricultural Drought, Dynamic Risk Assessment

Abstract: With the increasing impact of natural disasters, people have gradually begun to pay attention to disaster risk and incorporate risk management into various disaster reduction actions to reduce disaster losses more effectively. In order to solve the shortcomings of the existing research on dynamic assessment of agricultural drought risk, based on the discussion of the steps of dynamic assessment of drought risk and the components of drought risk, as well as BP neural network and random forest algorithm, this paper briefly discusses the indicators and weights of drought risk assessment and the establishment of empirical data samples, and discusses the design of dynamic assessment model of agricultural drought risk. Finally, the random forest evaluation model designed in this paper is compared with BP neural network and variable fuzzy controller. The experimental results show that the accuracy of random forests in assessing the regional risk level of drought is 97.3% on average. The recognition accuracy of random forest and BP neural network is better than that of variable fuzzy controller. Therefore, it is verified that the agricultural drought risk dynamic assessment method based on machine learning has high practical value.

1. Introduction

Dynamic assessment of agricultural drought risk. The distribution trend of drought risk can be obtained, and the distribution of drought resistance capacity can be further obtained, so as to provide a basis for drought resistance and disaster reduction, help reduce drought losses, and at the same time, the limited drought resistance can be invested in areas with weak drought resistance, so as to improve the effectiveness of drought resistance.

Nowadays, more and more scholars have done a lot of research in the dynamic assessment of agricultural drought risk through various technologies and system tools, and have also made certain research achievements through practical research. Baumhauer selected Gannan Prefecture as the research area, combined with GIS spatial analysis and statistical methods, and selected SPEI as the best time scale for drought monitoring of alpine meadow in Gannan Prefecture. In addition, based on SPEI index and drought risk probability model, drought ecological risk assessment was carried out. From the perspective of spatial distribution, in the past 10 years, drought in alpine meadow growing season mainly occurred in the east of Gannan Prefecture, while the frequency of drought in the west was relatively low [1]. RafiU used Copula function to calculate the joint recurrence time of moderate and severe drought in Weihe River basin of China, and detected the dynamic characteristics of drought and other multi-attribute risks through a sliding window with three year intervals. The main driving forces of drought dynamic evolution were studied by using cross wavelet transform. The results show that since the 1980s, the multi attribute risk of drought in the basin is generally high. The drought risk in the basin is closely related to sunspots [2]. Roy B D developed a new method for drought risk assessment system based on Hilbert Huang transform (HHT) method to understand the spatial distribution of groundwater drought risk. We used HHT to analyze the drought vulnerability, the temporal and spatial variation characteristics of groundwater level and the physical mechanism of groundwater factors, to quantify the sensitivity of groundwater to the environment, and to show the recovery capacity of each region. In addition, the drought hazard is determined using a standardized precipitation index, and the dynamic drought intensity gives the durability characteristics of each region. Drought exposure was also investigated to quantify livelihoods that meet people's water needs at specific population densities. The research results can be used for water resource management, drought resistance, sustainable development of groundwater resources and decision-making to determine drought risk [3]. Although the existing research on agricultural drought risk dynamic assessment method is very rich, there are still many problems in its practical application.

This paper mainly evaluates and analyzes the four major indicators, namely, the risk of disaster causing factors, the sensitivity of disaster pregnant environment, the vulnerability of disaster bearing bodies, and the ability to prevent and reduce disasters. The corresponding impact factors are evaluated and analyzed through the drought risk index method. Finally, a comprehensive weighted analysis is conducted to regionalize the drought disasters in Gansu Province. Based on this, the dynamic assessment model of agricultural drought risk is established through the random forest in the machine learning algorithm, in which the five fold cross method is used to train and verify the correctness of the model calculation results. Finally, by selecting BP neural network and variable fuzzy controller to compare the accuracy with the random forest evaluation model proposed in this paper, the experimental data verify that the evaluation effect of machine learning algorithm is better than the traditional evaluation model.

2. Research on Dynamic Assessment Design of Agricultural Drought Risk Integrated with Machine Learning

2.1. Dynamic Assessment of Drought Risk

According to the idea of drought risk assessment, the main steps of comprehensive assessment of regional drought risk are:

(1) Data collection: The collected data shall include the following contents: first, the regional background and basic information of disaster bearers in the region shall be assessed; second, the characteristics of drought disaster causing factors; third, the vulnerability of disaster bearers and the regional drought resistance and mitigation capacity; fourth, the historical disaster information [4].

(2) Risk identification: drought risk identification generally includes three aspects: identification of risk sources, identification of disaster bearing bodies and identification of disaster pregnant

environment [5].

(3) Assessment of drought risk degree: under the guidance of the systematic view of drought disaster, the risk assessment of drought disaster is to analyze and predict various risk indicators of drought disaster pregnant environment and disaster causing factors based on historical drought data [6].

(4) Assessment of vulnerability of hazard bearing bodies: generally, according to different types of hazard bearing bodies, the evaluation index system is constructed from the physical exposure of regional hazard bearing bodies, disaster loss sensitivity and regional drought resistance and disaster reduction ability, and then the analytic hierarchy process (AHP), fuzzy comprehensive evaluation and other methods are used for coupling to conduct comprehensive evaluation [7].

2.2. Drought Risk Components

(1) Risk: For drought risk, its risk refers to the probability of encountering drought of a certain degree in a certain regional environment at a certain time, which can be quantitatively described by the frequency or recurrence period in probability statistics [8].

(2) Disaster loss sensitivity: refers to the degree of vulnerability of a disaster bearing body to damage and damage caused by a certain disaster, which reflects the inherent attributes of a disaster bearing body under the action of different disaster causing factors [9]. For the drought risk, the higher the disaster loss sensitivity of the general bearing body is, the more vulnerable it is to the drought loss, and the higher the corresponding drought risk is [10].

(3) Drought resistance: refers to the ability of disaster bearing bodies to take corresponding adjustment measures to mitigate or avoid drought losses when facing drought [11]. Generally speaking, the higher the drought resistance capacity of the disaster bearing body, the smaller the corresponding drought risk. For example, in a drought prone area, the more drought resistant water conservancy projects are.

2.3. Machine Learning

(1) BP neural network

From the overall view of the neural network structure, it mainly has two modules: feature extraction module and classification module. The feature extraction module mainly extracts the features of the input image through its unique convolution structure to form a feature map. In the classification module, the input of the first full connection layer is the feature image obtained after feature extraction by the convolution layer and pooling layer [12].

(2) Random forest

A random forest is a large tree classifier consisting of a series of small tree branches. Each small branch is an independent tree. Each branch randomly selects a small group of input variables to be divided on each node, uses the CART method to plant trees, and does not prune them to maximize their size [13].

The theoretical basis is as follows: Set the classifier set as $[k_1(G), k_2(G), ..., k_i(G)]$, randomly extract the training set from the distribution of random vectors G and N, and define the total boundary function as:

$$mf(G,N) = cx_i L(k_i(G) = N) - \max_{s \neq N} cx_i L(k_i(G) = s) \tag{1}$$

Where, is the index function, $mf(G,N)$ is the difference measurement value, random vectors

G, N, and the degree to which the number of correct classifications exceeds the number of other classifications. The greater the difference, the better the classification effect [14].

Based on the large number theorem, the generalized error function WR^* is introduced to prove that the random forest converges everywhere and will not produce transition fitting. The calculation formula is as follows:

$$WR^* = W_{G,N}(mf(G,M) \le 0) \tag{2}$$

$$W_{G,N}(W_\alpha(k(G)=N) - \max_{s \ne N} W_\alpha(k(G,\alpha)=s) \le 0) \tag{3}$$

The upper bound of the generalized error function is expressed by the strength $f = Q_{g,n}ma(G,N)$ and the average correlation $\bar{\lambda} = \dfrac{Q_{\alpha,\alpha}(\lambda(\alpha,\alpha))sd(\alpha)sd(\alpha^{'})}{Q_{\alpha,\alpha}(sd(\alpha)sd(\alpha^{'}))}$ [15].

3. Research on Dynamic Assessment of Agricultural Drought Risk by Integrating Machine Learning

3.1. Drought Risk Assessment Indicators and Weights

In this paper, 13 small factors under these four indicators are finally determined by reading a large number of references, drawing on domestic and foreign research methods and combining the drought characteristics and actual conditions of Gansu Province [16]. The weight value of each index factor is determined by using the analytic hierarchy process. Table 1 shows the drought risk assessment indicators and weight values of Gansu Province. The weighted comprehensive analysis method combined with the natural breakpoint method was used for classification, and the spatial distribution map was obtained, which was finally used for the study of drought risk assessment [17].

Table 1. Drought risk assessment indicators and weights in Gansu province

Target layer	Criterion layer	Weight value	Indicator layer	Weight value
Risk assessment	Risk of disaster causing factors	0.35	Light drought	0.1
			Moderate drought	0.2
			Severe drought	0.3
			Extreme drought	0.4
	Susceptibility of pregnant environment	0.25	Percentage of precipitation anomaly	0.5
			Vegetation coverage	0.3
			Field capacity	0.2
			Population density	0.4
	Vulnerability of disaster bearing body	0.3	Economic density	0.2

			Proportion of cultivated land area	0.3
			Land average large livestock	0.2
	Ability to prevent and mitigate disasters	0.1	Revenue	0.6
			Area affected by drought and flood	0.3

3.2. Test Data Sample

After completing the data preparation, the vulnerability assessment of different years is carried out by building a model. According to the statistical analysis, as of 2019, 621 of 4541 assessment units in the study area had experienced disasters; By 2020, 589 regions have suffered disasters; As of 2021, a total of 214 assessment units have experienced landslide disasters. It can be seen that drought stricken areas basically account for about 1/10 of the whole study area [18]. According to the data structure distribution, there are also some differences in the sampling proportion. The sample construction is shown in Table 2 below:

Table 2. Model sample construction

Particular year	Split sample set	Drought stricken area	Other areas	Sampling proportion
2019	Train	621	2147	10%
	Verification	118	1254	
2020	Train	589	2457	12%
	Verification	127	1056	
2021	Train	214	987	13%
	Verification	98	567	

4. Application Research on Dynamic Assessment of Agricultural Drought Risk Integrated with Machine Learning

4.1. Dynamic Assessment of Agricultural Drought Risk Integrated with Machine Learning

Using the established training set and verification set samples, the five fold cross method is used to train the verification model, that is, the random sampling with placement is set by the disaster risk level. The trained model was used to assess the risk of 4527 drought stricken Fengxian District in Gansu Province. Its modeling flow chart is shown in Figure 1.

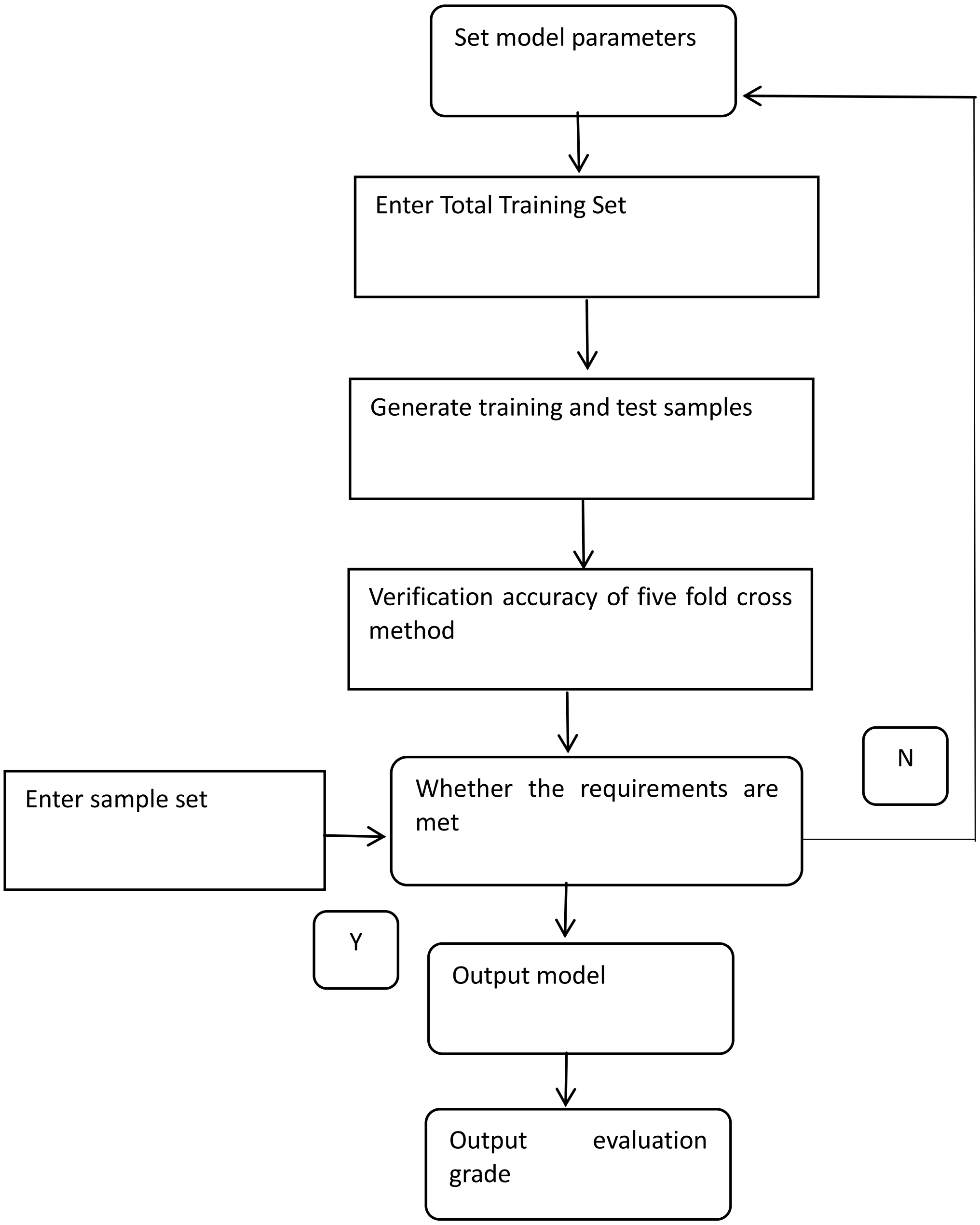

Figure 1. Flow chart of random forest drought risk assessment

Random forest modeling includes two processes: classification tree growth, voting and model prediction. The modeling process is as follows:

(1) There are n indicators, and each indicator has m samples. The total sample size is S, the test set is G, the verification set is R, the number of classification trees is k, the depth of the tree is x, the number of node characteristics is f, and the node termination condition is the minimum number of samples or the minimum information gain.

(2) Bagging method is used to randomly select r times for putting back, construct a new training set, and start training as the root node.

(3) Each training set is the input data of a single classification tree. If the node reaches the termination condition during training, the node is a leaf node. When the predicted output value of

the node is the maximum value of the tree, the category recording the maximum value accounts for the proportion of the sample set; If the termination is not reached, y indicators are randomly selected from n indicators to search for the threshold with the best classification effect.

(4) Repeat step (3) until all nodes of each tree have been trained or reach the termination condition.

(5) According to the generated multiple classifiers, predict the output of the data. The classification result is determined by the number of votes. The model with the highest number of votes is the required classification model.

4.2. Application of Dynamic Assessment of Agricultural Drought Risk Integrated with Machine Learning

Use spatial analysis tools to calculate the classification grade of agricultural drought at historical points and the classification accuracy of the model, analyze the random forest classification evaluation model proposed in this paper, compare the evaluation accuracy with BP neural network and variable fuzzy controller, and the algorithm calculation results are shown in Table 3.

Table 3. Algorithm evaluation results

Algorithm	Variable fuzzier	BP neural network	Random forest
Low risk	56.2%	91.8%	95.4%
Medium risk	76.5%	92.7%	96.9%
high-risk	79.9%	93.8%	98.7%
Very high risk	77.3%	95.7%	98.5%

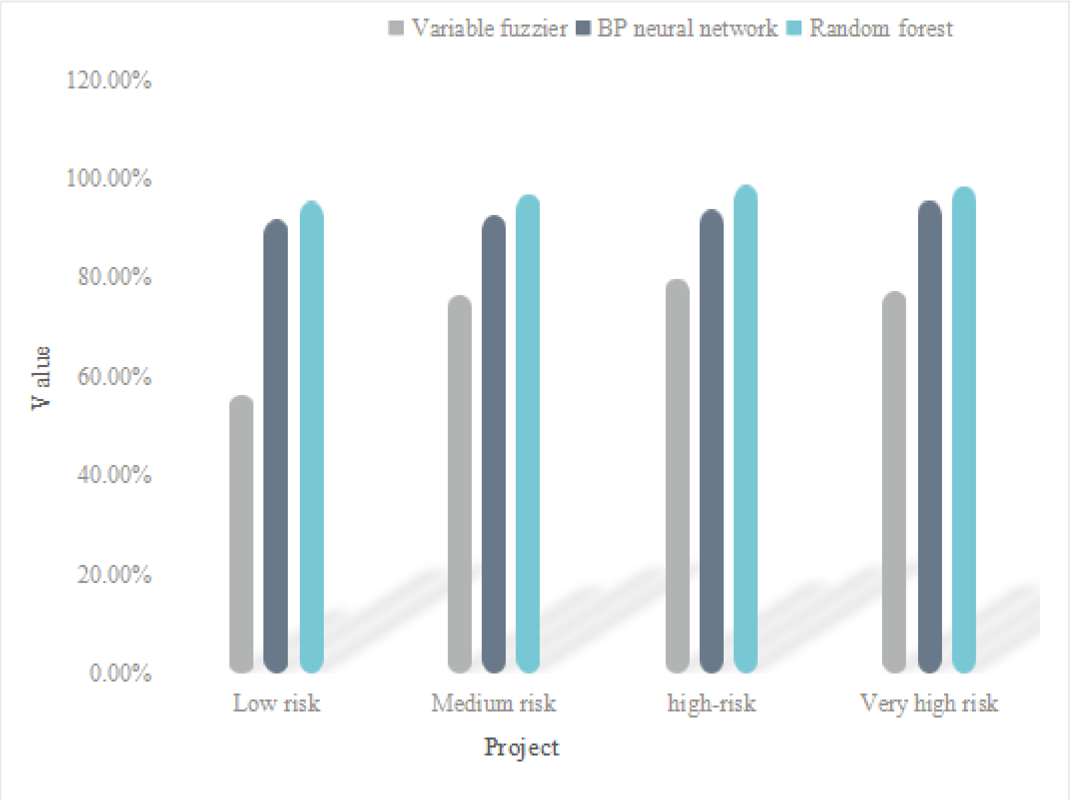

Figure 2. Comparison of algorithm evaluation performance

It can be seen from the data in Figure 2 that the accuracy of the variable fuzzy identification model in the extremely high risk area is 77.3%, the accuracy in the high risk area is high, and the accuracy in the low risk area is low, only 56.2%; The accuracy of BP neural network model in extremely high risk area is 95.7%, that in high risk area is 93.8%, and that in low risk area is 91.8%; The accuracy of high risk areas of random forests is 98.5%, and the distribution of each grade is good, among which the accuracy of high risk areas is 98.7%. The accuracy of the machine learning model is better than that of the variable fuzzy recognition. The machine learning model evaluates the risk level of mountain torrents by learning the historical mountain torrent disaster information, which has obvious advantages over the traditional mathematical method in judging the risk level by index information.

5. Conclusion

In this paper, through the data collection and extraction of disaster data, the drought disaster database of Gansu Province is constructed, and the temporal and spatial distribution law and occurrence mechanism of historical drought disasters in Gansu Province are analyzed. Select drought disaster risk assessment indicators according to disaster analysis, use the random forest model to complete the construction of drought disaster risk assessment model in Gansu Province, and use the drought disaster database data to verify the accuracy of the model. The experimental results show that the random forest algorithm is better than the BP neural network algorithm, indicating that the integrated classifier is more comprehensive in the learning process, more accurate in extracting historical characteristic values, more powerful in data mining, and better in classification and prediction, The weak classifier has some shortcomings, such as over fitting, which can be avoided by the integrated classifier. By using machine learning to assess the risk level, the commonly used method of dividing county boundaries is broken, which makes the risk level more objective and fair.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Baumhauer, Judith, Mitten, et al. Using PROs and machine learning to identify "at risk" patients for musculoskeletal injury. Quality of life research: An international journal of quality of life aspects of treatment, care and rehabilitation, 2018, 27(Suppl.1):S9-S9.

[2] Rafi U, Ahmad S, Zia A, et al. Contribution of apolipoprotein estimations to cardiovascular risk assessment in women with endometriosis. Journal of biological regulators and homeostatic agents, 2021, 35(3):1195-1198.

[3] Roy B D, Jiang Y, Graham C, et al. A Structured Benefit-Risk Assessment to Evaluate a Combination of Olanzapine and Samidorphan for the Treatment of Schizophrenia and Bipolar I Disorder. CNS Spectrums, 2022, 27(2):233-233. https://doi.org/10.1017/S1092852922000311

[4] Muhammad, Bilal, Azmi, et al. Risk assessment of insulin resistance with anthropometric measures in non-diabetic heart failure patients: A gender based case-control investigation at tertiary care hospitals of Karachi, Pakistan. Pakistan journal of pharmaceutical sciences, 2019, 32(1(Supplementary)):255-260.

[5] Sd A, It B. Interpretable machine learning approach in estimating traffic volume on low-volume roadways - ScienceDirect. International Journal of Transportation Science and Technology, 2020, 9(1):76-88. https://doi.org/10.1016/j.ijtst.2019.09.004

[6] Shetty C, Sowmya B J, Seema S, et al. Air pollution control model using machine learning and IoT techniques - ScienceDirect. Advances in Computers, 2020, 117(1):187-218. https://doi.org/10.1016/bs.adcom.2019.10.006

[7] Sombolestan S M, Rasooli A, Khodaygan S. Optimal path-planning for mobile robots to find a hidden target in an unknown environment based on machine learning. Journal of ambient intelligence and humanized computing, 2019, 10(5):1841-1850. https://doi.org/10.1007/s12652-018-0777-4

[8] Chittora D. How Al and Machine Learning Helps In Up Shilling to Better Career Opportunities. Pc Quest, 2019, 32(3):20-21.

[9] Diederichs, M. S. Early assessment of dynamic rupture hazard for rockburst risk management in deep tunnel projects.. Journal of the Southern African Institute of Mining and Metallurgy, 2018, 118(3):193-204. https://doi.org/10.17159/2411-9717/2018/v118n3a1

[10] Tadepalli S K, Lakshmi P V. Application of Machine Learning and Artificial Intelligence Techniques for IVF Analysis and Prediction. International Journal of Big Data and Analytics in Healthcare, 2019, 4(2):21-33. https://doi.org/10.4018/IJBDAH.2019070102

[11] Karimanzira D, Rauschenbach T. Performance Prediction of a Reverse Osmosis Desalination System Using Machine Learning. Journal of Geoscience and Environment Protection, 2021, 9(7):16. https://doi.org/10.4236/gep.2021.97004

[12] Lopes R G, Ladeira M, Carvalho R N. Use of machine learning techniques in the prediction of credit recovery. Advances in Science Technology and Engineering Systems Journal, 2017, 2(3):1432-1442. https://doi.org/10.25046/aj0203179

[13] Oosterhoff J H F, Karhade A V, Oberai T, et al. Prediction of Postoperative Delirium in Geriatric Hip Fracture Patients: A Clinical Prediction Model Using Machine Learning Algorithms:. Geriatric Orthopaedic Surgery & Rehabilitation, 2021, 12(2):435-446. https://doi.org/10.1177/21514593211062277

[14] Le T, Bondarev A V, Kozlova E V. Application of Machine Learning Algorithms in Predicting Pyrolytic Analysis Result. IOP Conference Series: Earth and Environmental Science, 2021, 931(1):012013 (10pp). https://doi.org/10.1088/1755-1315/931/1/012013

[15] Masud F A, Rejaul M, Royel I, et al. Smart Risk Prediction Tools of Appendicitis Patients: A Machine Learning Approach. Biointerface Research in Applied Chemistry, 2021, 11(1):7804-7813. https://doi.org/10.33263/BRIAC111.78047813

[16] Andrew, Tzer-Yeu, Chen, et al. Convolutional neural network acceleration with hardware/software co-design. Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies, 2018, 48(5):1288-1301.

[17] Lima R, Suriamin F, Marfurt K, et al. Convolutional Neural Networks. AAPG Explorer, 2018, 39(10):22-23.

[18] *Subject-Independent Emotion Recognition of EEG Signals Based on Dynamic Empirical Convolutional Neural Network. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2021, 18(5):1710-1721. https://doi.org/10.1109/TCBB.2020.3018137*