Journal of Computers Vol. 33 No. 5, October 2022, pp. 39-47
doi: 10.53106/199115992022103305004

# Predicting Credit Accessibility Based on Social Capital and Artificial Intelligence

Thi Thu Huong Dinh[1,2], Thai Thuy Lam Ha[3*], Hongjuan Yang[1]

[1] Faculty of Management and Economics, Kunming University of Science and Technology, Kunming, 650093, China
maihuong.ng96@gmail.com, 939738277@qq.com
[2] Le Quy Don Technical University, Hanoi, 100000, Viet Nam
[3] Education Department, Dong Thap University, Dong Thap, 870000, Vietnam
httlam@dthu.edu.vn

Received 9 November 2021; Revised 2 February 2022; Accepted 3 March 2022

**Abstract.** Social capital is a significant factor that determines the success of a business, company or individual. The paper presents a new method to predict the aquaculture household's ability to access to credit, based on social capital and artificial intelligence. Social capital data is collected through surveys of aquaculture households, and then replying on artificial intelligence algorithms to build predictive models. The prediction results show that our proposed method can accurately predict households accessing the formal credit market. This is the basis for the planning and development of aquaculture in potential areas.

**Keywords:** social capital, credit accessibility, artificial intelligence

## 1 Introduction

Aquaculture sector in Vietnam has a long history and is developing strongly, typically raising fish, shrimp, frogs, etc., contributing to creating jobs, increasing incomes and reducing poverty [1-2]. However, due to the rapid conversion, a large area of rice, cropland, and low-yielding fields is facing many difficulties and challenges. In addition to facing difficult problems such as natural conditions, environment, diseases, natural disasters, asynchronous planning and development, aquaculture households also face difficulties in accessing input and output resources. Especially, market information is disturbed leading to many difficulties in accessing official credit sources [3].

Social capital has been discussed by many researchers over the years and there have been many studies related to this issue. Many studies have showed that social capital has significant impacts on individuals, businesses, organizations, etc. [4-6]. The presence of social capital increases farmers' ability to adopt new technologies [7]. For rural finance and credit, there have been many studies indicating the important role of social capital in accessing credit of individuals [8-9]. Heikkila et al. [10] did a survey in Uganda to study the link between social capital and financial access. The results showed a positive association between social capital and access to credit, the clear impact of personal social capital on the poor. Besides, it also found that personal social capital promotes access to semi-formal and informal financial institutions. Alam et al. [11] studied the factors and social capital affecting the ability to access to credit and technology in rural Bangladeshi households. This research indicated that this access is significantly influenced by social capital and institutions. To help poor household better cope with extreme natural conditions, government interventions through planned adaptation are important. Linh et al. [12] summarized the characteristics of the rural credit market in Vietnam, the determinants of credit accessibility and the comparison with developed countries. The results of the study indicated the characteristics and credit accessibility in Vietnam with the participation and intervention of the government. The impacts of credit on output, household income and poverty reduction are also highlighted in this study. Lu et al. [13] pointed out the impacts of social capital on bank lending. The results showed that households with a higher level of social capital are less likely to be rejected during the loan application process. In addition, households with a higher level of social capital will bring a higher rate of return after getting a bank loan.

To assess the impacts of social capital on credit accessibility, most studies use quantitative methods such as Logistic, Profit, Tobit etc. and multivariate regression to build models [14-16]. These algorithms have good results for linear data, but for non-linear data, there can be large errors. In recent years, artificial intelligence has been greatly developed thanks to its accuracy and learning ability. Currently, it is applied in many fields such as information processing, language, images, etc. [17-19]. In economics, artificial intelligence is applied to build mathematical models for data about growth, financial performance, supply, demand, etc. [20-22]. Janh [23] used

* Corresponding Author

artificial neural network to predict gross domestic product growth. The results showed that the artificial neural network can predict more accurately than the linear network, thereby confirming that the artificial neural network is suitable for the statistical model. Meng [24] based on the agricultural indicators of China's rural areas and fuzzy min-max neural network to suggest economic practices for each region and necessary policies to promote the sustainable agricultural development.. Andres et al. [25] used multilayer artificial neural network to predict economic index for developing economies. This study filled the gap caused by the lack of data, from which the government can provide effective policies. Compared to traditional methods, predicted results are superior. International economic agencies and organizations will benefit from this method.

To increase the predictability of Vietnamese aquaculture households' credit accessibility. This study uses artificial intelligence and social capital to build a predictive model of access to credit. Social capital data is collected from a survey of aquaculture households in the north of Vietnam. Matlab is used as the environment to build models. The effectiveness of the proposed method is evaluated through experiments.

## 2 Materials and Methods

### 2.1 Data Collection and Processing

As a basis for research, data have been collected from aquaculture households in the Red River Delta of Vietnam including Hung Yen, Thai Binh, Nam Dinh, Hai Duong, and Hai Phong provinces. Information and data collection have been done by direct interview with the aquaculture farmers, and through trained collaborators, as well as under the supervision of the research team. We have used fractional sampling in combination with random sampling, which helps to reduce inequality problems and increase representation of areas with similar characteristics. In addition, we have also conducted interviews with several households with long-term experience and experts to assess, review the current situation and gain a deeper understanding of the actual situation in the data collection area. Finally, we have selected the input data to build an analytical model including approaching fisheries extension association, union organization, farm management board, credit officer, output, colleagues and friends, age of household heads, number of years in occupation, number of years living in the locality, qualification of household heads.

To increase the processing performance and accuracy of machine learning algorithms, we need to normalize the data [26]. There are many main methods to normalize data, such as min-max normalization or z-score normalization. In this study, we have used the max normalization method, choosing the largest number to normalize the remaining data.

### 2.2 Random Forest Algorithm

Random forest (RF) algorithm was proposed by Breman [27], it is an ensemble learning algorithm. RF combines classification, regression tree and bagging algorithm. In the training process, random feature selection is introduced to reduce the correlation of samples, thereby solving the overfitting problem of a single decision tree model, and making the model have good noise tolerance. The bagging algorithm uses a random sampling method to sample $m$ rounds of the original data set to obtain a new data set. The decision tree learning process adopts a top-down recursive method and uses information entropy as a measure to construct a tree with the fastest decrease in entropy. Classification and regression trees use the Gini coefficient as the basis for feature splitting. Suppose we have K categories, and the probability of the $k$-th category is $p_k$, then the Gini coefficient of the probability distribution is shown in Equation (1).

$$\text{Gini} = \sum_{k=1}^{K} p_k\left(1 - p_k\right) = 1 - \sum_{k=1}^{K} p_k^2 \ . \tag{1}$$

Given sample set $Q$, its Gini coefficient is shown in Equation (2).

$$\text{Gini}(Q) = 1 - \sum_{k=1}^{K}\left(\frac{|C_k|}{|Q|}\right)^2 \ . \tag{2}$$

where, $C_k$ is the sample subset belonging to the *k*-th category in *Q*.

*Q* can be divided into two parts $Q_1$ and $Q_2$ according to feature *A*.

$$Q_1 = \{(x, y) \in Q | \ A(x) = a\} . \quad \textbf{(3)}$$

$$Q_2 = Q - Q_1 . \quad \textbf{(4)}$$

where, $(x, y)$ is represents a sample in the sample set *Q*, *x* is the eigenvalue of the sample, *y* is the label of the sample; a is a possible value. Then under the condition of feature *A*, the Gini coefficient of sample set *Q* is defined as Equation (5).

$$\text{Gini}(Q, A) = \frac{|Q_1|}{|Q|}\text{Gini}(Q_1) + \frac{|Q_2|}{|Q|}\text{Gini}(Q_2) . \quad \textbf{(5)}$$

### 2.3 Whale Swarm Algorithm

Whale swarm algorithm (WSA) is a swarm intelligence optimization algorithm that simulates the predation behavior of humpback whales in the ocean [28]. The whales attacked through the bubble net, continuously spiraling upward to approach the target, gradually shrinking the encircling circle, and finally reached the target location. The algorithm mainly consists of 3 stages: foraging, bubble-net preying and searching.

*Foraging*

In the process of searching for food, WSA makes the current optimal solution or near optimal solution as the food target. The other whales move towards the closest food at present and gradually shrink the whole enclosure. The the process as shown in Equation (6).

$$X_{i+1} = X_i - A\left|CX_i^* - X_i\right| . \quad \textbf{(6)}$$

where $X_i^*$ is the current optimal position of the all swarms, $X_i$ is the current location, i is the iteration, $X_{i+1}$ is the next location; *A* and *C* are calculated by (7, 8), a is a constant, $t_1$ and $t_2$ are random numbers in [0,1].

$$A = 2at_1 - a . \quad \textbf{(7)}$$

$$C = 2t_2 . \quad \textbf{(8)}$$

*Bubble-net preying*

The bubble-net preying process to find local optimization, the mathematical of this process is Equation (9).

$$X_{i+1} = \begin{cases} X_i - A\left|CX_i^* - X_i\right|, & p < 0.5 \\ \left|X_i^* - X_i\right| e^{bl} \cos(2\pi l) + X_i^*, & p \geq 0.5 \end{cases} . \quad \textbf{(9)}$$

where *l* and *p* are random number, *b* is a constant.

*Searching*

The searching process to find global optimization. Each whale has a random location, and it is shown in Equation (10).

$$X_{i+1} = X_{\text{rand}} - A\left|CX_{\text{rand}} - X_i\right| . \quad \textbf{(10)}$$

where $X_{\text{rand}}$ is the random location.

### 2.4 Build an Analysis Model

In this study, we use RF and WSA algorithms to build an analytical model of credit market access. The accuracy of the RF algorithm is mainly affected by the randomly selected feature number $m$ and the decision tree $n$. The size of $m$ is connected to the decision tree's performance as well as the correlation between decision trees. The smaller the $m$, the weaker the ability of a single decision tree, and there will be a tendency to reduce the performance of the RF algorithm; nevertheless, the lower the correlation between the decision trees, the more powerful the classifier. The size of $n$ determines the convergence speed and accuracy of the algorithm. Experiments have proved that it is not that the larger $n$ is, the higher the accuracy. Normally, there are a pair of $m$, $n$ parameter combinations with the best accuracy within a specific threshold, or find the parameter with the highest accuracy within the allowed range of operating efficiency. Therefore, one of the most difficult problems in random forest is determining how to choose the parameters $m$ and $n$. In this research, we apply WSA to solve this problem and determine the optimal $m$ and $n$ combination. The combination of RF and WSA is shown in Fig. 1. The WSA-RF algorithm is the name given to this combo.

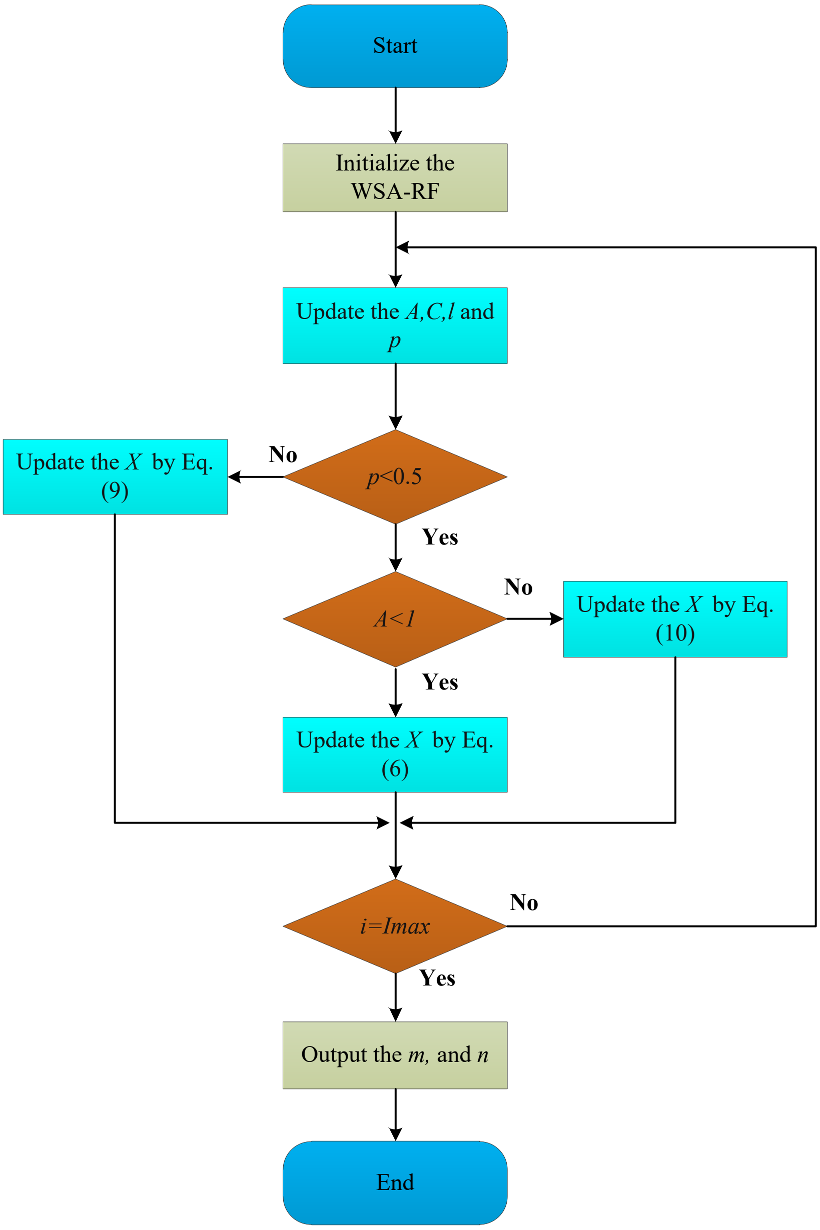


**Fig. 1.** Flow chart of WSA-RF algorithm

**Algorithm 1.** WSA-RF algorithm

```
Initialize the WSA-RF
for i = 1: Imax
   for j = 1: N
      Update the a, A, C using Equations (7 and 8)
      if p < 0.5
        if A < 1
          Update the X by Equation (6)
         else
          Update the X by Equation (10)
        end if
       else
    Update the X by Equation (9)
      end if
     end for
end for
```

## 3. Results

Through a survey of 467 aquaculture households, there are 355 households that need loans from different credit sources, accounting for 76% of the total surveyed households. Out of these 355 households, 242 have access to formal credit markets, while the rest have to borrow from other sources of credit. This shows that formal credit accessibility of aquaculture households has some shortcomings.

The data are divided into two categories. The first category is households that have access to the formal credit market and the second category is the households that cannot borrow from the formal credit market. Based on survey data and the proposed algorithm, we build a model to predict formal credit accessibility. Fig. 2 shows the training process to build and optimize the model. We select the number of swarms to be 50 and the number of iterations to be 100. The results show that the model quickly converges; by the 19th iteration, it has reached the maximum optimal level, and the deviation has decreased from 3.4% to 0.8%. The optimization process takes place within 74s. The WSA algorithm can quickly optimize the model, increasing the predictability of the model. WSA is a reliable algorithm for parameter optimization. Using WSA to optimize the parameters of the RF algorithm helps to improve the accuracy of RF and giving reliable results.

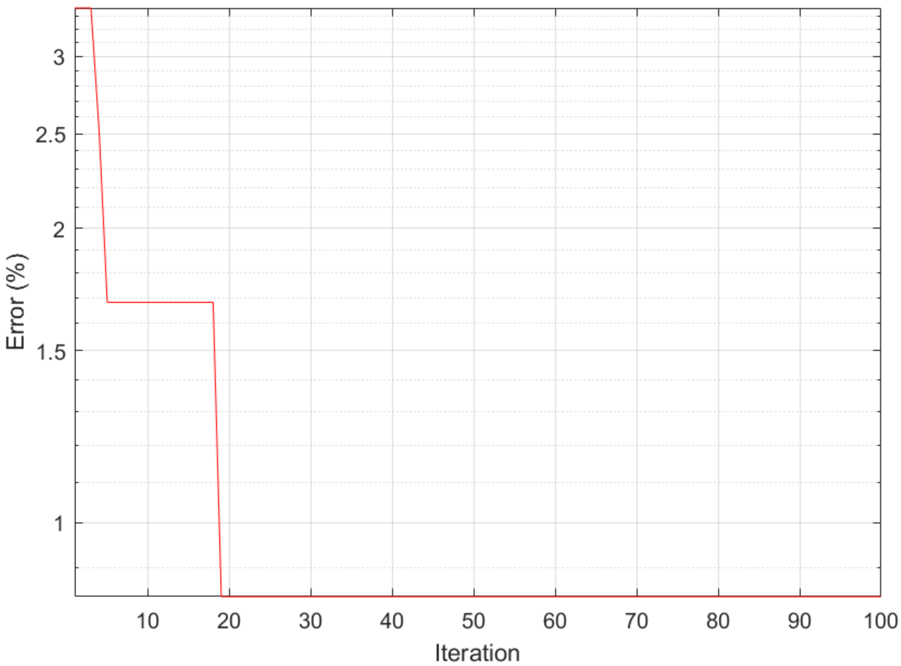


**Fig. 2.** The credit accessibility prediction model's training and optimization process

**Table 1.** The accuracy of the credit accessibility prediction model

| Class | Accuracy (%) |
|---|---|
| 1 | 99.6 |
| 2 | 98.2 |
| Overall | 99.2 |

The prediction results are shown in Fig. 3. Here, the first class we label as 1, and the second class we label as 2. We can see that the difference between the predicted point and the actual point is very small. Specifically, there is only a 3 point difference between class 1 and class 2. It can be seen that the model has very good predictive results. Table 1 is the accuracy of the predicting credit accessibility model for aquaculture households. It can be seen that the accuracy of the first class is 99.6%, that of the second class is 98.2% and the overall accuracy is 99.2%. This result proves that our proposed model has good robustness and fitting performance. The model can predict the credit borrowing capacity of aquaculture households well, serve as an important basis for households with loan needs. At the same time, it serves as a foundation for local governments to design particular aquaculture development strategies.

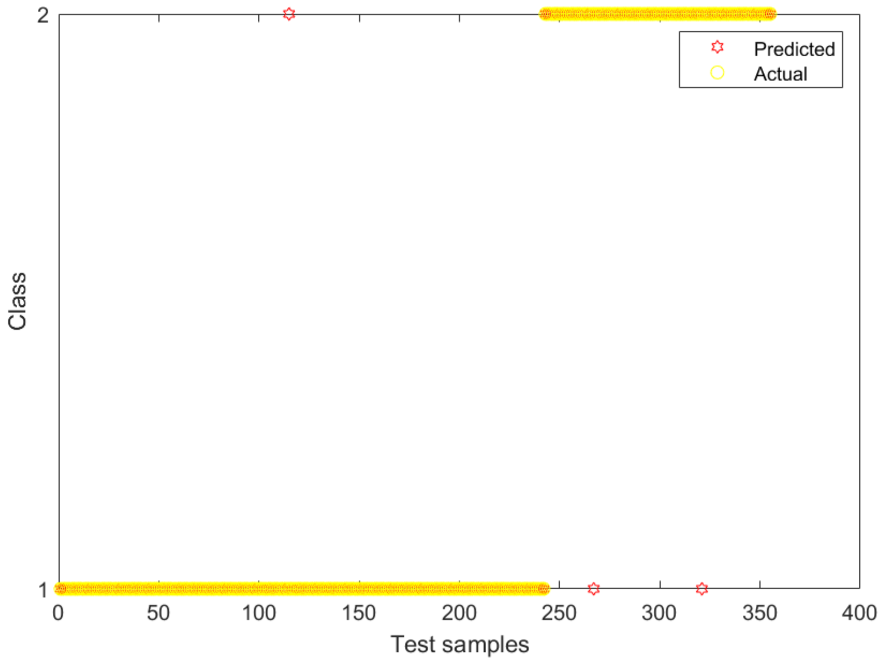


**Fig. 3.** The results of predicting credit accessibility of the WSA-RF model

## 4. Discussion

### 4.1 The Positive Characteristics Affecting Access to Credit

Through the optimization process of the model, we find that the characteristics of the social capital have a great impact on the ability to access formal credit. These characteristics include the fisheries extension association, farm management board, credit officers, friends, and colleagues, which are also mentioned in the study of Rahman et al. [29] hay của Duong [3]. The fisheries extension association has a positive impact on formal credit accessibility and is significant in the model settings. If the household head has access to the fisheries extension association, the rate of access to formal credit will increase by 10%.

The farm management board has a positive impact on formal credit accessibility. When aquaculture households receive the support, advice, sharing and help of the aquaculture area' the management board on the issue of accessing official credit, it will increase their ability to get loans.

Credit officers also play an important role in aquaculture households' access to formal credit. The help, advice and sharing of the credit officer will help the householder quickly access the credit market.

Colleagues and friends also positively affect the ability to get formal credit loans of aquaculture households. If shrimp farmers receive more help and share credit experiences from friends, they will increase their ability to access formal credit.

Thus, there are four important factors that increase formal credit accessibility of aquaculture households. Among these, the credit officer plays the most important role so that aquaculture households can obtain formal credit.

### 4.2 Comparison of Different Methods

In this section, we will compare the proposed method with some traditional prediction methods. These methods include logistic regression (LR) [30], artificial neural networks (ANN) [31-32], and support vector machines (SVM) [33]. The comparison results are shown in Table 2 and Fig. 4. We can see that the proposed method has a higher accuracy than the other methods. Specifically, the method in this study is 4% higher than the LR method, 2.5% higher than the ANN method and 4.8% higher than the SVM method. The LR and SVM algorithms are based on statistical probability theory. LR uses Sigmoid function as a non-linear factor, thereby performing binary classification problems. The SVM algorithm uses the kernel function as a non-linear factor to perform multi-classification. The ANN algorithm uses a neural network that simulates the human nervous system, through linear or non-linear transformations to classify the data features. The algorithm in this study uses the ensemble learning method combined with a optimization problem, thereby increasing the classification ability of the model. This result proves that our proposed method is superior, suitable for the problem of predicting the access to credit market of aquaculture households in Vietnam.

**Table 2.** The accuracy of different methods

| Methods | Accuracy (%) |
|---|---|
| LR | 95.2 |
| ANN | 96.3 |
| SVM | 94.4 |
| WSA-RF (This study) | 99.2 |

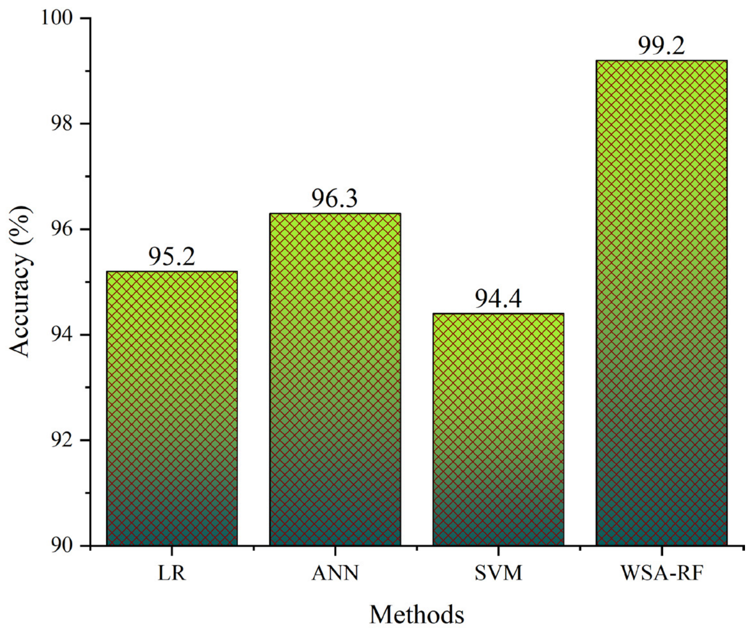


**Fig. 4.** Comparison of accuracy of different methods

## 5. Conclusion

In this study, we propose a method to predict credit accessibility based on artificial intelligence algorithms and social capital. The data is taken from the actual survey on the status of official loans of aquaculture households in Vietnam. The results of our model have accurately predicted the ability to access the credit market. Compared with traditional methods, our method achieves higher predictive results. This is the basis for developing aquaculture in Vietnam.

## References


[1] O.M. Joffre, J.R. De Vries, L. Klerkx, P.M. Poortvliet, Why are cluster farmers adopting more aquaculture technologies and practices? The role of trust and interaction within shrimp farmers' networks in the Mekong Delta, Vietnam, Aquaculture 523(2020) 735181.

[2] A.T.N. Hai, S. Speelman, Economic-environmental trade-offs in marine aquaculture: The case of lobster farming in Vietnam, Aquaculture 516(2020) 734593.

[3] T.D. Duong, The impact of social capital on the activities of shrimp farmers in the Mekong Delta, [doctoral thesis] Ho Chi Minh City: Ho Chi Minh National University, 2020.

[4] E.C. Flores, D.C. Fuhr, A.M. Bayer, A.G. Lescano, N. Thorogood, V. Simms, Mental health impact of social capital interventions: a systematic review, Social psychiatry and psychiatric epidemiology 53(2)(2018) 107-119.

[5] M. Dumas, J.L. Barker, E.A. Power, When does reputation lie? Dynamic feedbacks between costly signals, social capital and social prominence, Philosophical Transactions of the Royal Society B 376(1838)(2021) 20200298.

[6] X. Neumeyer, S.C. Santos, A. Caetano, P. Kalbfleisch, Entrepreneurship ecosystems and women entrepreneurs: A social capital and network approach, Small Business Economics 53(2)(2019) 475-489.

[7] A.S.S. Ville, G.M. Hickey, U. Locher, L.E. Phillip, Exploring the role of social capital in influencing knowledge flows and innovation in smallholder farming communities in the Caribbean, Food Security 8(3)(2016) 535-549.

[8] Q.H. Dinh, T.B. Dufhues, G. Buchenrieder, Do connections matter? Individual social capital and credit constraints in Vietnam, The European Journal of Development Research 24(3)(2012) 337-358.

[9] T. Pham, O. Talavera, Discrimination, social capital, and financial constraints: The case of Viet Nam, World Development 102(2018) 228-242.

[10]A. Heikkila, P. Kalmi, O.P. Ruuskanen, Social capital and access to credit: Evidence from Uganda, The Journal of Development Studies 52(9)(2016) 1273-1288.

[11]G.M. Alam, K. Alam, S. Mushtaq, Influence of institutional access and social capital on adaptation decision: Empirical evidence from hazard-prone rural households in Bangladesh, Ecological Economics 130(2016) 243-251.

[12]T.N. Linh, H.T. Long, L.V. Chi, L.T. Tam, P. Lebailly, Access to rural credit markets in developing countries, the case of Vietnam: A literature review, Sustainability 11(5)(2019) 1468.

[13]H. Lu, B. Wang, H. Wang, T. Zhao, Does social capital matter for peer-to-peer-lending? Empirical evidence, Pacific-Basin Finance Journal 61(2020) 101338.

[14]M.M. Hasan, A. Habib, Social capital and payout policies, Journal of Contemporary Accounting & Economics 16(1) (2020) 100183.

[15]Z. Liu, Y. Zhang, H. Li, Digital Inclusive Finance, Multidimensional Education, and Farmers' Entrepreneurial Behavior, Mathematical Problems in Engineering 2021(2021).

[16]B. Li, Y. Shen, Effects of land transfer quality on the application of organic fertilizer by large-scale farmers in China, Land Use Policy 100(2021) 105124.

[17]Y.S. Mohamed, H.M. Shehata, M. Abdellatif, T.H. Awad, Steel crack depth estimation based on 2D images using artificial neural networks, Alexandria Engineering Journal 58(4)(2019) 1167-1174.

[18]C.T. Jensen, X. Liu, E.P. Tamm, A.G. Chandler, J. Sun, A.C. Morani, N.A. Wagner-Bartak, Image quality assessment of abdominal CT by use of new deep learning image reconstruction: initial experience, American Journal of Roentgenology 215(1)(2020) 50-57.

[19]R.W. Filice, Deep-learning language-modeling approach for automated, personalized, and iterative radiology-pathology correlation, Journal of the American College of Radiology 16(9)(2019) 1286-1291.

[20]L. Liu, X. Zhan, Analysis of financing efficiency of Chinese agricultural listed companies based on machine learning, Complexity 2019(5)(2019) 1-11.

[21]M.H. Rafiei, H. Adeli, Novel machine-learning model for estimating construction costs considering economic variables and indexes, Journal of construction engineering and management 144(12)(2018) 04018106.

[22]X. Wei, W. Chen, X. Li, Exploring the financial indicators to improve the pattern recognition of economic data based on machine learning, Neural Computing and Applications 33(2)(2021) 723-737.

[23]M. Jahn, Artificial neural network regression models: Predicting GDP growth (No. 185), HWWI Research Paper (2018).

[24]X. Meng, M. Liu, M. Wang, J. Wang, Q. Wu, Fuzzy min-max neural network with fuzzy lattice inclusion measure for agricultural circular economy region division in Heilongjiang province in China, IEEE Access 8(2020) 36120-36130.

[25]A.R. Andres, A. Otero, V.H. Amavilah, Using deep learning neural networks to predict the knowledge economy index for developing and emerging economies, Expert Systems with Applications 184(2021) 115514.

[26]R. Vaddi, P. Manoharan, Hyperspectral image classification using CNN with spectral and spatial features integration, Infrared Physics & Technology 107(2020) 103296.

[27]L. Breiman, Random forests, Machine learning 45(1)(2001) 5-32.
[28]S. Mirjalili, A. Lewis, The whale optimization algorithm, Advances in engineering software 95(2016) 51-67.
[29]T. Rahman, S. Ara, N.A. Khan, Agro-information Service and Information-seeking Behaviour of Small-scale Farmers in Rural Bangladesh, Asia-Pacific Journal of Rural Development 30(1-2)(2020) 175-194.
[30]F.S. De Menezes, G.R. Liska, M.A. Cirillo, M.J.F. Vivanco, Data classification with binary response through the Boosting algorithm and logistic regression, Expert Systems with Applications 69(2017) 62-73.
[31]G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70(1-3)(2006) 489-501.
[32]B.S. Raghuwanshi, S. Shukla, Minimum class variance class-specific extreme learning machine for imbalanced classification, Expert Systems with Applications 178(2021) 114994.
[33]C. Zhou, J.G. Chase, G.W. Rodgers, Support vector machines for automated modelling of nonlinear structures using health monitoring results, Mechanical Systems and Signal Processing 149(2021) 107201.