Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 3741087, 7 pages
https://doi.org/10.1155/2022/3741087

Research Article

Application of Big Data Analysis Techniques in Sports Training and Physical Fitness Analysis

Xinwen Li,[1] Xiaowei Chen,[2] Lihui Guo ,[3] and Christine A. Rochester[4]

[1] *Department of Physical Education, University of Electronic Science and Technology of China, Chengdu Sichuan 610054, China*
[2] *Department of Athletics and Swimming, Chengdu Sport University, Chengdu, Sichuan 610000, China*
[3] *The Graduate School, Chengdu Sport University, Chengdu, Sichuan 610000, China*
[4] *Department of Exercise Science, Physical Education and Recreation of Colorado State University-Pueblo, Pueblo, Colorado 81001, USA*

Correspondence should be addressed to Lihui Guo; guolihui2022@163.com

Received 20 January 2022; Revised 21 February 2022; Accepted 24 February 2022; Published 25 March 2022

Academic Editor: Kalidoss Rajakani

With the development of sports and information technology, people use mathematical tools and computer technology to study sports data and mine the intrinsic value of sports data. Statistical methods are the most widely used to achieve this goal. The research purpose of sports effect evaluation research is to understand the impact of sports on physical fitness through mining and analysis of sports data and to provide theoretical guidance for the public to participate in fitness activities scientifically and effectively. At present, in the study of combining individual performance test data, the research on the standardization of physical fitness monitoring data for sports training is relatively scarce. Therefore, based on the background of big data, this paper integrates the existing data standardization work and designs a plan for the standardization of physical fitness monitoring data for sports training. Combined with machine learning, data preprocessing is performed to obtain the data required by the machine model. The comprehensive physical fitness rating model and the recommendation model are established to realize the development of physical fitness monitoring service applications. In the experiment, compared with the three classical methods, the results show that the classification accuracy of this paper is 4% higher than that of other algorithms, which can more intuitively reflect the characteristic samples of sports training. In this paper, the data mining and analysis technology based on feature indicators in the mining and application of sports data has great application value for human fitness guidance and has certain research value and market application prospects.

1. Introduction

Sports data is an important part of big data resources. Mining and analysis of sports data can effectively understand the impact of sports on the human body and exercise efficacy. Existing sports data mining methods mainly focus on extracting and constructing effective basic sports data features and use statistical methods to analyze sports data in combination with basic features [1]. However, with the rapid development of technologies such as data mining and machine learning, the mining of sports data cannot be simply carried out using statistical methods. How to effectively mine and analyze sports data by combining machine learning and data mining technology so that it can provide useful suggestions for mass physical exercise is an urgent problem that needs to be studied. Sports data mining is an important direction and application of big data analysis [2].

In the "Current Situation, Problems and Thinking of Physical Fitness Monitoring at the Grassroots Level," it is pointed out that the main problem in physical fitness monitoring of sports training is that, due to the lack of professionals in grass-roots monitoring work and the prolonged testing time, this may lead to unreliable test data [3]. Therefore, the physical fitness monitoring work should implement a combination of sports and medicine. The relevant operations of this measure are as follows: the establishment of personal files can further achieve the effectiveness of data preservation; promoting the convenience of information

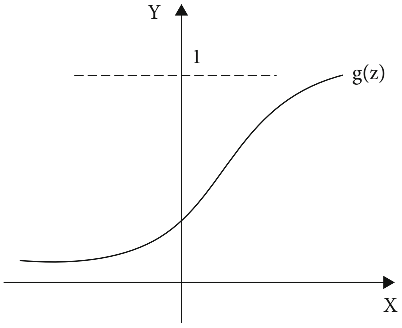

Figure 1: Function value.

tracking can realize data sharing and exchange, so that it can obtain self-training information based on personal basic data [4]. Hongyan adopted the decision tree ID3 algorithm and association rule Apriori algorithm for data mining analysis. Through the analysis of the ID3 algorithm, she got what factors were related to motor function. Through the analysis of the Apriori algorithm of association rules, the influence of sports training on the body was excavated [5]. The physical health test data of Caihong's research was studied from a deeper level by using the FP-Growth algorithm. Through the results of algorithm operation, it was observed that there was a lack of lower limb strength training in sports training. The vital capacity level and endurance level were obviously weak, so it was recommended to strengthen the training of aerobic exercise [6].

This paper integrates machine learning into sports training fitness and establishes an evaluation model. Based on the background of big data, the test information of physical fitness monitoring of sports training and physical fitness is used as sample data. Machine learning is used to establish a physical fitness evaluation model for sports training, which includes data preprocessing, model establishment, and evaluation so that it can achieve effective data integration. The standardization of the data file storage format can realize the upload of sports training physical fitness test data files so that it can solve the work efficiency problem of data processing and improve the work efficiency of sports training physical fitness monitoring data processing.

The main innovations of this paper are as follows:

(1) A scheme for the standardization of sports training physical fitness monitoring data is proposed to solve the problem of heterogeneous formats of monitoring service application platforms

(2) The combination of machine learning and data processing improves the work efficiency of comprehensive evaluation in sports training physical fitness monitoring

(3) By combining machine learning and the test data information of sports training physical fitness, it provides a reference for the recommended methods of fitness methods

2. Related Work

2.1. The Basic Principle of the Decision Tree. Regarding the machine models studied, they are similar to the decision tree classification model, which is equivalent to a function value obtained by combining multiple two-class problem models, as shown in Figure 1.

The task of the computer is to determine what is in the current picture. After the computer makes relevant judgments, it is necessary to test whether the judgment output results meet the expected effect [7]. Therefore, the feature vector X can be set in the computer to represent the result. Then, the current computer can use the computer professional language $Y = 0$ or $Y = 1$ to express the right or wrong of the judgment result, and the main formulas used are shown in

$$h_0(x) = g\left(0^T x\right), \tag{1}$$

$$g_0(z) = \frac{1}{1 + e^{-z}}. \tag{2}$$

From formulas (1) and (2) and Figure 1, it can be concluded that the classification judgment $g(z)$ can consider that the function value of its output correlates with a certain interval threshold. Next, z is transposed into a linear problem, and the value of z can be calculated according to the position of the point (below or above the line); the result can be obtained [8]. Furthermore, the specific value of z can also be obtained according to the graph of $g(z)$. For example, when the threshold value $g(z) = 1$ and according to Figure 1, it can be known as follows.

$$\begin{cases} z < 0, g(z) < 0.5, \\ z = 0, g(z) = 0.5, \\ z > 0, g(z) > 0.5. \end{cases} \tag{3}$$

2.2. Evaluation Principle of the Classification Model. In machine learning, machine learning is used to solve binary classification problems. Compared with reality, the solution to the problem will inevitably have a certain deviation. Therefore, a conceptual evaluation of the performance of the machine model is performed [9].

2.2.1. Related Concepts of TP, TN, FP, and FN. When classifying in the machine model, the meanings of TP and TN both represent the situation of the classification result: TP is the positive class and TN is the negative class. The meaning of FP means that the wrong category is divided into right, and FN means that the correct category is divided into wrong.

2.2.2. Precision, Recall, Accuracy, and H-Mean. Precision is correctly predicted as a positive class, which refers to the proportion of the positive class division in all predictions, as shown in

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{4}$$

Recall is also correctly predicted as a positive class, which refers to the proportion of positive class divisions in all positive classes, as shown in

$$\text{Re call} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (5)$$

Accuracy refers to the proportion of correct predictions (positive and negative) in all predictions, as shown in

$$H\text{-mean} = \frac{2\text{TP}}{2\text{TP} = \text{FP} = \text{FN}}. \quad (6)$$

H-mean ($F1$) evolved from the relevant formula. When the $F1$ value decreases, the TP increases relatively, while the error class decreases relatively. Therefore, both precision and recall increase relatively; that is, $F1$ weights both precision and recall, as shown in

$$H\text{-mean} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}. \quad (7)$$

2.2.3. Preprocessing of Big Data Analysis. With big data as the background, machine learning is used to process, analyze, and clean the test information of physical fitness monitoring of sports training. The collection of preprocessed physical fitness test data is used as sample data for the establishment and evaluation of physical fitness evaluation models [10].

Min–max normalization, also known as dispersion normalization, is a linear transformation of the original data, so that the resulting values are mapped between $[0, 1]$, as shown in

$$x^* = \frac{x - \min}{\max - \min}. \quad (8)$$

In formula (8), max represents the maximum value of the sample and min represents the minimum value of the sample.

Based on the mean and standard deviation of the original data, the standardization of the data is completed. The preprocessed data conforms to the standard normal distribution; that is, the mean is 0 and the standard deviation is 1, as shown in

$$x^* = \frac{x - H}{\theta}. \quad (9)$$

In formula (9), H represents the mean of the total sample and θ represents the standard deviation of the total sample.

3. Establishment of the Physical Fitness Evaluation Model for Sports Training

Before the establishment of the physical fitness evaluation model for sports training, first, the ratings of excellent, good, qualified, and unqualified are used in the sports training fitness model, and the labels are set for each data separately. Then, the decision tree algorithm in machine learning is applied to build the model, and the main role is to solve the classification problem.

3.1. Evaluation Process of Physical Fitness in Sports Training. The physical fitness evaluation model of sports training is established, and its process is shown in Figure 2.

It can be seen from Figure 2 that the main steps of constructing a physical fitness evaluation model for sports training are as follows.

(1) The basic information included in the model is recorded as follows: selecting response time, vertical jump, standing on one foot, height, weight, vital capacity, step index, sitting forward flexion, grip strength, push-up, which are used as a single inner node

(2) To find the optimal node, traversal is used to find it

(3) After step 2, two different leaf nodes can be branched and generated

(4) This model will automatically repeat steps 2 and 3, the final learning results are traversed and obtained. This result is used as the root node of the evaluation model to satisfy the physical fitness of sports training and completes the establishment of the physical fitness evaluation model of sports training [11]

For the root node searched above, it is necessary to determine the quality of the effect. Therefore, in the process of automatic learning of the machine model, in order to determine the ability of the machine model to automatically select the root node, it is necessary to use the gained information to determine whether the root node is the optimal node. The specific algorithm is shown in

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{\text{Values}(A)}^{A} \text{Entropy}(S). \quad (10)$$

In formula (10), entropy (S) stands for information entropy, which means the purity of the sample set selected in this study. The smaller its value, the higher the purity of its sample set. S represents the total number of records in all root nodes, and A represents the value of each attribute in the sample data set.

The algorithm and formula will take the maximum information gain as the test result of the root node. However, to reduce the sample attributes with a larger number of attributes, formula (11) is introduced.

$$\text{GainRatio}(s, A) = \frac{\text{Gain}(s, A)}{\text{IntrinsicValue}(A)}. \quad (11)$$

3.2. Generation of an Evaluation Model for Physical Fitness in Sports Training. The evaluation model of physical fitness of sports training only needs to process the information of individual physical fitness test, analyze the feature importance of the sports training physical fitness evaluation

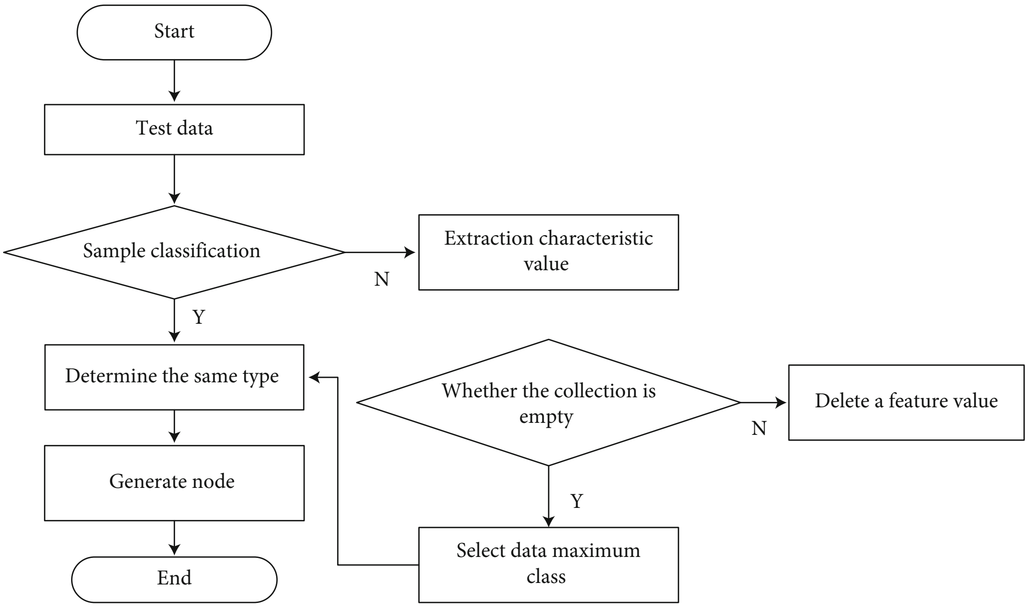

Figure 2: The process of the comprehensive evaluation model.

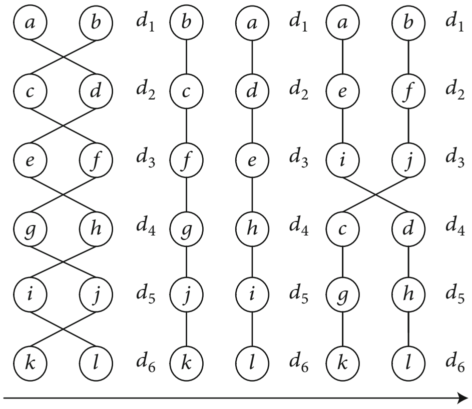

Figure 3: The driven model of evaluation.

model, and find the root node by combining the information gain. For the decision tree of the physical fitness evaluation model for exercise training in this paper, the fitness test data set will select a certain fitness test index feature to calculate the information gain and Gini coefficient and use this index as the optimal feature of impurity. The driving information of the evaluation model is shown in Figure 3.

In Figure 3, the evaluation model of physical fitness of sports training takes $tjzs$ (step index) as the root node and takes $tjzs$ ($w = 3.5$) as the judgment condition, so that the Gini coefficient of the model can be calculated as 0.476. Meanwhile, the 77 physical fitness test sample data will be divided into 4 types, the sample data sets are 0, 0, 30, and 47, and the data set with the most types is unqualified. Taking it as the root node, the decision tree gradually grows, and the next layer of recommendation nodes for exercise training physical fitness evaluation is generated, in which the left node selects fhl (lung capacity) as the optimal feature and takes fhl (lung capacity) $W4.5$ as the judgment condition, the Gini coefficient is 0.355, and the sample data sets are 0, 0, 12, and 40, respectively. The right node selects fhl (lung capacity) as the optimal feature and takes fhl (lung capacity) £1.5 as the judgment condition, and the Gini coefficient is 0.403. The 25 physical fitness test sample data are also divided into 4 types, and the sample data sets are 0, 0, 18, and 7, respectively. The data set with the most types is qualified. Until the Gini coefficient of the physical fitness test data set is zero, the decision tree belonging to the evaluation model of sports training fitness will stop growing. Compared with the physical fitness test data set of the root node, the impurity of the remaining branch nodes is relatively reduced. It shows that the growth direction of the decision tree is conducive to the division of physical fitness evaluation of sports training.

4. Experimental Analysis

4.1. Experimental Setup and Evaluation Criteria. The data used in the experiment is the sports database SED. Experiments are carried out on the SED database, which is set as training set and test set with a ratio of 4 : 1. The preprocessed physical fitness test data set (12143 pieces of physical fitness test sample data) is applied to the research on the physical fitness evaluation model of sports training, and the establishment of the model is realized. Among them, 70% of the physical fitness test sample data (8500 pieces) are directly used for model building and training; the remaining 30% of the physical fitness test sample data (3643 pieces) are applied to evaluate the performance of the model. Therefore, this paper mainly studies the influence of physical exercise on the changes of the body's indicators. The data obtained from the sports exercise experiment is used as the positive class, and the nonsports data is used as the negative class, and a comparative experiment is carried out on the two kinds of data.

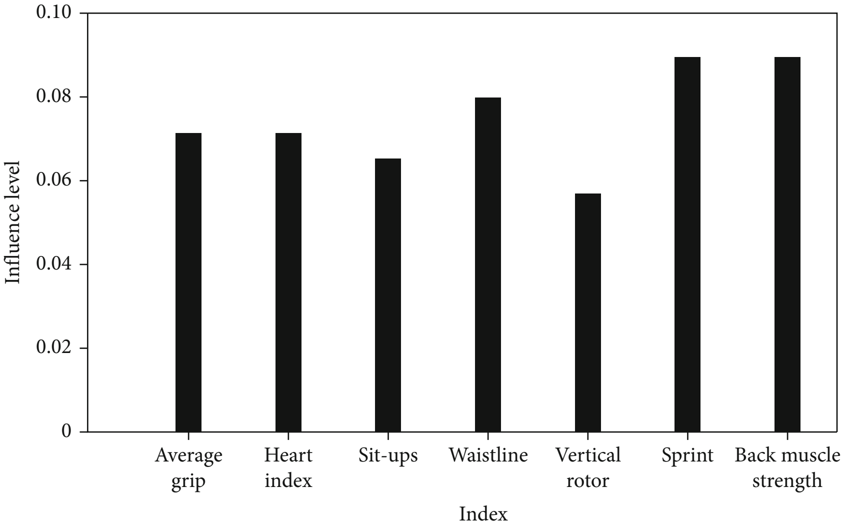

Figure 4: The degree of influence of wrestling sports indicators.

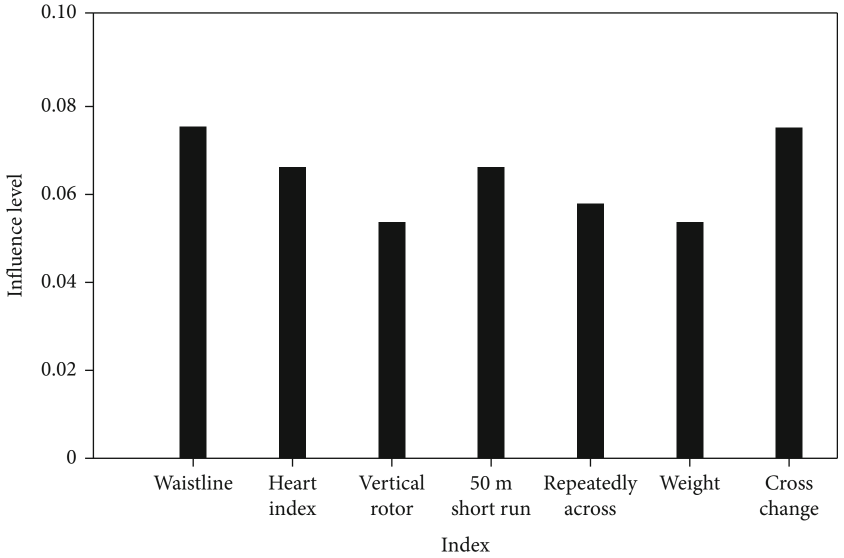

Figure 5: The degree of influence of competitive sports indicators.

Table 1: Top-k accuracy comparison results of methods on SED.

Method	Wrestling			Competitive		
	Top-3	Top-5	Top-10	Top-3	Top-5	Top-10
Mean	0.33	0.4	0.7	0	0.2	0.4
Variance	0.33	0.2	0.7	0	0	0.4
RFE	0	0.2	0.3	0	0.2	0.4
SVM	0	0.4	0.6	0.33	0.6	0.7
Ours	0.66	0.4	0.6	0.33	0.6	0.7

The experimental evaluation criterion in this paper is the accuracy of top-k, which is defined as the ratio based on the algorithm-obtained body metric that matches the ground truth. The higher the accuracy, the more effective the algorithm is. The formula for calculating the accuracy is as follows.

$$\text{Precision} = \frac{k}{n}, \tag{12}$$

where k represents the number of impact indicators consistent with the ground truth and n represents the total number of indicators selected in the ground truth.

4.2. Experimental Results. In this paper, the importance of the influence of folk sports on physical indicators based on feature gain is ranked.

Figure 4 shows the ranking of the effects of wrestling sports on body indicators. It can be seen that the most influential are the average grip strength, cardiac function index,

Table 2: Comparison of evaluation results.

Index	Unnormalized data	Min–max data
Accuracy	0.91542	0.91885
$F1$	0.82253	0.81014
Precision	0.72255	0.75253
Recall	0.72656	0.72656

one-minute sit-ups, waist circumference, standing turns, 50 m sprint, and back muscle strength index.

Figure 5 shows the ranking of the impact of competitive foot sports on physical indicators. It can be seen that the most influential factors are waist circumference, cardiac function index, standing body rotation, 50 m sprint, repeated traverse, weight, and cross-direction running. Competitive foot sports mainly exercise physical function and reaction ability, especially the physical function of the lower body.

The experiments in this paper are compared with the methods using mean, variance, support vector machine (SVM), and recursive feature elimination (RFE). The training set and test set are set in the same way. The experimental results are shown in Table 1.

The accuracy of the evaluation of the physical impact indicators obtained by the method in this paper for the two types of sports data is the highest on top-5. For further experiments with different accuracy parameter settings, the results of top-3 and top-10 can be acquired. The accuracy rates obtained by the method in this paper are significantly higher than those obtained by other methods, which proves that the method in this paper is more effective for evaluating the impact of sports on physical indicators.

4.3. Experimental Evaluation. This paper uses 30% of the sample data (3643 items) to evaluate the model performance. After the normalization of data, the performance evaluation model of the physical fitness for sports training is evaluated, and the comparison results are obtained, as shown in Table 2.

It can be seen from Table 2 that after data preprocessing using min–max standardization for the physical fitness test data, the corresponding index values of the three evaluation indicators before and after are as follows: the accuracy rate of 90.55% and 90.52%, the H-mean value of 81.23% and 79.64%, the accuracy rate of 83.06% and 79.13%, and recall rates of 79.84% and 80.19%, respectively.

5. Conclusion

This paper focuses on the mining and analysis of exercise data and analyzes the characteristics of big data to predict and analyze the impact of exercise on human performance. The main work and steps are as follows:

(1) Based on sports data, combined with feature selection and data mining theory, a more effective sports data mining method is proposed

(2) It can be seen from the experimental results that different types of sports have different exercise effects and different physical indicators

(3) From the experimental results, we can also see that different types of sports have different exercise effects and different physical indicators

Through the use of the results obtained by the assessment method, the physical training is guided accordingly, and the physical exercise is strengthened purposefully.

In the future research, the recommendation of sports training methods and the algorithm and model of physical fitness monitoring can be further improved. Using a variety of algorithms to mine the value of physical fitness test data can promote the model to expand in breadth. In addition, with the normalization of information sharing, follow-up related research can also be closely linked with multiplatform-related data so as to achieve the increase in different types of data. Further in-depth mining of data value can also promote the extension of the model in depth.

Data Availability

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

Conflicts of Interest

The authors declare that there is no conflict of interest with any financial organizations regarding the material reported in this manuscript.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities of China: Physical Education and Health Promotion (No. ZYGX2018007FRJH), and the Educational Reform Project of University of Electronic Science and Technology of China: to construct a customized system of campus physical health and create a new model of physical education (No. 2015XJYYB080).

References

[1] R. P. Bonidia, J. D. Brancher, and R. M. Busto, "Data mining in sports: a systematic review," *IEEE Latin America Transactions*, vol. 16, no. 1, pp. 232–239, 2018.

[2] Z. ZhiBin and L. Ying, "Collaborative innovation research on national physique monitoring and university student physique health test," *Advances in Physical Sciences*, vol. 7, no. 4, pp. 156–161, 2019.

[3] Z. Tong, L. Xiaofeng, and Z. He, "Med-PPPHIS: blockchain-based personal healthcare information system for national physique monitoring and scientific exercise guiding," *Journal of Medical Systems*, vol. 43, no. 9, pp. 305–323, 2019.

[4] B. Xu, D. Huang, and B. Mi, "Research on E-commerce transaction payment system basedf on C4. 5 decision tree data mining algorithm," *Computers, Networks & Communications*, vol. 35, no. 2, pp. 113–121, 2020.

[5] A. Saxena, N. Brault, and S. Rashid, *Big Data and Artificial Intelligence for Healthcare Applications*, CRC Press, 2021.

[6] A. S. Tewari and A. G. Barman, "Sequencing of items in personalized recommendations using multiple recommendation techniques," *Expert Systems with Applications*, vol. 97, pp. 70–82, 2018.

[7] M. I. S. Saidin, "The Arab Spring through Malaysian youth 'eyes': knowledge, perceptions and influences," *Mediterranean Journal of Social Sciences*, vol. 9, no. 1, pp. 121–135, 2018.

[8] I. Ahmed, S. Obermeier, S. Sudhakaran, and V. Roussev, "Programmable logic controller forensics," *IEEE Security and Privacy*, vol. 15, no. 6, pp. 18–24, 2017.

[9] J. Bispo and J. M. P. Cardoso, "A MATLAB subset to C compiler targeting embedded systems," *Software: Practice and Experience*, vol. 47, no. 2, pp. 249–272, 2017.

[10] M. B. S. Ahmad and A. Cheung, "Automatically leveraging MapReduce frameworks for data-intensive applications," in *2018 International Conference on Management of Data*, Houston TX USA, 2018.

[11] Z. Lijie, "Evaluation research on data processing of mental health of college students based on decision tree algorithm," *Journal of Computational Methods in Science and Engineering*, vol. 19, no. 4, pp. 1101–1108, 2019.