Physical exercise is very necessary for students, only good physical quality can better study. The mining of physical exercise data is the analysis and integration of physical exercise data. The purpose of this paper is to study the data mining technology of students’ physical exercise. This paper proposes a data mining algorithm for student physical exercise based on Cloud Computing (CdC). In the traditional data mining technology, data regression and clustering are added, and the data is processed through variance and covariance, which can make the data complete the processing steps such as cleaning faster and saves a lot of data processing time. The experiments in this paper show that the algorithm proposed in this paper can analyze the physical exercise data very well. For data samples with sample sizes of 1000 and 2000, the running time is about 20 seconds lower than the original algorithm, and the resource factor of the system is stable at about 82%, which proves that the algorithm has a good load balance.

1. Introduction

With the rapid development of information technology and storage technology, more and more computers are used in various fields, and a large amount of historical data has been accumulated in various fields over time. And over time, these skyrocketing volumes of data have reached staggering numbers. Usually, due to the large amount of data, traditional statistical methods or simple data analysis methods cannot obtain useful information from these massive data. Sometimes, even if the data sets are relatively small, traditional processing methods are powerless for these data due to some nontraditional characteristics of the data itself. In addition, there are some problems that cannot be solved using existing data analysis techniques. But there is a lot of information hidden in these data that is difficult for people to know, but it has very important information for some applications in real life. This hidden information can often be used to guide decision-making and business conduct in many areas. If these data are not used simply because the means of analyzing data are backward, then the data will lose the opportunity to be revisited and utilized, which will be a great waste of data resources. In this case, it is necessary to develop methods that can process massive data sets and technologies to mine and analyze information. Data mining is produced in this situation. Data mining methods can discover information and laws that cannot be found by traditional data processing methods. Because of the effectiveness and efficiency of data mining, data mining technology has been widely used. Physical exercise, as the most important form of mass physical activity, can enhance body health and fitness level, improve physical appearance, regulate mental and emotional health, improve social relations, and even improve the overall national physique. Physical exercise covers a wide range of people, and the forms, contents, and methods are more flexible and diverse. Formally, an individual can choose to exercise alone or with friends. The content includes entertainment, combat, health and corrective sports and fitness, and bodybuilding. Exercise methods are divided into transformation method, repetition method, competition method, etc., which are commonly used in daily teaching and training, as well as some exercise
methods created in practice, such as morning exercise and interclass exercise. According to different age, gender, occupation, and health status, exercisers should carry out appropriate exercise activities according to their own physical conditions and participate in physical exercise in a targeted, step-by-step, and persistent manner. In the process of physical exercise, it is necessary to establish correct exercise values, and avoid problems that cause negative emotions to be vented through exercise, and that the load exceeds the range of physical functions, which is detrimental to physical and mental health. In addition, if it can combine movement and stillness, form and spirit, and internal and external during physical exercise and apply traditional Chinese body training methods to it, the benefits for exercisers will be even greater. At the same time as the physical fitness is improved, the mind is also calmed, which is helpful for meditation.

College students are the pillars of national construction and the driving force for the sustainable development of schools. College students must develop in an all-round way. In the comprehensive quality of all students’ development, sports development is also very important. The level of physical quality reflects an important standard for how much a college student can participate in future social activities. The educational philosophy of the school is very important to the development of college students. In university courses, the form, content, and innovative teaching methods of physical education are the keys to the development of physical education in colleges and universities. The current state of college students’ physical fitness is generally not ideal, especially in some colleges and universities, where students only perform limited physical exercise for physical fitness tests. This thinking mode is similar to the traditional Chinese test-taking mode. That is, to exercise physical fitness to pass the exam and not make sports a part of life. Therefore, it is necessary to use the data mining research of student physical exercise based on CdC to illustrate whether the exercise level of the student group has reached the standard.

This article has the following two main innovations in the writing of the article: (1) CdC and data mining technology are applied to the research on physical exercise. In today’s digital world, everything can be quantified. Through the powerful data analysis ability of CdC and the data analysis ability of data mining technology, it can effectively analyze the physical exercise of students. (2) It proposes a data mining algorithm for physical exercise based on CdC. This kind of algorithm is targeted, and it can mine and analyze the physical exercise time, time period, and the basic data of students, and the results formed by such data mining technology are more convincing.

2. Related Work

We all know that the data contains a wealth of useful information. Because the size of data has increased by an order of magnitude since the invention of the computer, data mining research has become very popular. Helma et al. applied data mining technology to games, and their research model effectively simulates 2.5D fighting games [1]. Buczak et al. used literature analysis to conduct an analysis of data mining techniques to support network analysis machine learning for intrusion detection, and their findings provided some recommendations on when to use each method [2]. Xu et al. believes that data mining technology’s growing popularity and development has posed a serious threat to the security of personal sensitive data. To that end, they discuss the privacy issues that each type of user faces as well as the methods that can be used to protect sensitive data [3]. In a systematic review of the application of machine learning, data mining techniques, and tools in diabetes research, Kavakiotes et al. linked biotechnology and data mining techniques [4]. Chaurasia and Pal also conduct research on female breast cancer using biotechnology and data mining as a foundation. They used Weka software to compare three classification techniques, and the results show that Sequential Minimal Optimization (SMO) has a higher prediction accuracy than IBK and BF Tree methods, with Chaurasia V96.2 percent [5]. Yan and Zheng investigated the use of data mining technology in the stock market. They constructed a “universe” of more than 18,000 fundamental signals from financial statements and used a guided approach to assess the impact of data mining on fundamental-based anomalies. They found that many fundamental signals are significant predictors of cross-sectional stock returns even after accounting for data mining [6]. In terms of physical exercise research, more research is on the human body-related functions of exercise. Grassmann et al. believe that even a single physical exercise can enhance executive function. Therefore, they conducted a systematic review of the literature to analyze articles evaluating executive function in children with ADHD following acute exercise. They concluded that 30 minutes of physical activity could improve executive function in children with ADHD [7]. Håkansson and others concluded that the increase in BDNF they found after physical exercise is more likely to be of peripheral rather than of central origin. However, the association between BDNF levels and cognitive function after intervention may have implications for BDNF reactivity in serum as a potential marker of cognitive health [8].

It can be found that in the current research, there is almost no intersection between data mining and physical exercise. Although data mining technology has been used in many aspects, there is not much research on the analysis of physical exercise data.

3. CdC and Data Mining Technology

3.1. CdC. Cloud computing is the product of the integration of traditional computer technologies and network technologies such as grid computing, distributed computing, parallel computing, utility computing, network storage, virtualization, and load balancing. It aims to integrate multiple relatively low-cost computing resources into a resource pool with powerful computing and storage capabilities through the network and push this powerful computing capability to end-users with the help of advanced business models such as SaaS, PaaS, and IaaS.
3.2. Data Mining. The emergence of data mining (DM) technology [10] enables people to obtain meaningful information from massive data, and people can make decisions and make plans based on the obtained information. Data mining can extract information that is not known in advance but has potential value from a large amount of noisy data. Data mining is the core of database knowledge discovery (KDD) technology. Data mining technology originates from knowledge discovery and is an important step in knowledge discovery. Therefore, data mining is not only the mining of data but also the discovery and refinement of information, laws, knowledge, and other things that people are interested in and are useful for social activities.

3.2.1. Data Mining General Process. Figure 2 shows the following:

1. Determining the business objects: data mining can be carried out with goals, and the calculation ideas will be clearer.

2. Inputting data: input the raw data to be processed into the computer system. These raw data may be WEB data, sensor data, XML documents, or database data.

3. Data preprocessing: (1) data cleaning—remove the noise data mixed in the original data, accurately determine the abnormal points, and delete the vacant data. (2) Data integration, which integrates databases from multiple data sources. (3) Data selection—select data according to the following data mining requirements and ensure that the selected data can make the subsequent analysis work smoothly. (4) Data transformation, in order to make the mining results more in line with our requirements, the data form can be properly transformed.

4. Data mining: to obtain useful associations or patterns from the results of mining.

5. Postprocessing, including pattern assessment and knowledge targeting: (1) pattern evaluation—put the discovered patterns into practical applications to test the performance of the patterns and use trusted knowledge to check the contradictions that may exist in the patterns. (2) Knowledge identification—the results of data mining are represented by visual means and the results are displayed to customers.

6. Useful information: the final result of mining, which is generally useful information that people are interested in.

3.2.2. Data Mining and Classification in Cloud Environment

(1) Principle of Web Crawler. The purpose of web crawler is to automatically extract the content of web pages. The workflow of a web crawler is to start with one or several URLs, obtain the URL of this page, and then recursively read the content of the web page, so as to collect information related to the entire site, until it stops when certain conditions are met. The web crawler is also an important part of the search engine, and the overall structure is shown in Figure 3.

As shown in Figure 3, the web crawler crawls the information on the web in sections. The role of the repository is to store the web pages and URL links crawled by the crawler.
There are various methods of web crawler, and its principle is more complicated [11]. Most of the web crawler principles are based on the following steps: (1) obtaining the URL of the webpage content through the webpage analysis algorithm and putting it into the queue; (2) using the search method to pick the URL from the queue as the next object; and (3) repeating the second step until the URL is stopped when conditions such as depth restriction or area restriction occurs [12, 13].

(2) Python-Based Web Crawler. Commonly used languages such as Python, Java, and PHP can collect network data. Using crawler technology in daily life, Python has always been the most widely used language because of its concise language and easy-to-master, powerful HTTP library and HTML interpreter, good crawler framework Scrapy, and advantages in text character processing. More and more people start to use Python web crawler for web data mining [14].
(3) Hadoop and MapReduce Programming Model. Hadoop is maintained by the Apache Foundation, which allows users to realize distributed program development without knowing the underlying details of the distribution, and without having deep distributed development technology, an open source distributed system basic framework. The framework is hard-coded with reference to the related papers of Google’s Cdc troika, MapReduce, GFS, and Bigtable [15]. The Hadoop framework structure diagram is shown in Figure 4.

At the bottom of the distributed framework is Hadoop Distributed File System (HDFS), which corresponds to Google File System [16]. HDFS is a distributed file storage system. It provides features such as high reliability, load balancing, and redundant replication, ensuring the safe and reliable storage of massive data on storage media. At the same time, the throughput of I/O in the distributed file system is improved by means of redundancy, which provides the possibility for the realization of parallel computing of massive data. In massive data computing, the data transmission time is often longer than the actual computing time, so HDFS is the most basic factor to ensure the operating efficiency of the Hadoop framework. On top of HDFS is the MapReduce framework, and massive data parallel computing is implemented on the MapReduce programming framework. The data read and written during the running of the program are placed in HDFS, so that the application can effectively utilize the high I/O throughput of HDFS, thereby accelerating the overall running speed of the application [17].

MapReduce is a new distributed model abstracted from the features of functional programming. In terms of implementation, details such as parallelization, fault tolerance, data distribution, and balancing are hidden, and the entire distributed process is regarded as a function-like process expressed by map/reduce. It logically maps the input data into multiple key/value pairs and converts the key/value into key1/value1 through the corresponding function operation in the map. Then, in the reduce phase, the same key will be aggregated to achieve the final purpose of data manipulation. Therefore, in actual programming, developers only need to clarify the meaning of the key and value in each stage of data map/reduce, and then they can convert the existing business to the distributed parallel computing platform [18, 19].

Figure 5 depicts the MapReduce workflow. The Hadoop framework implements data splitting and merging using two classes of Mapper and Reducer. Mapper provides setup and cleanup to manage resources in the life cycle of Mapper, in addition to map functions for data segmentation. Setup is called after the Mapper is finished and before the map action is executed, so all of the data shared by each mapper that needs to be preread should be preread in the Setup. After all map actions have been completed, cleanup is called, and it is primarily used to clear temporary data. In practice, the map/reduce job divides the input data set into several independent data blocks, which are then processed in parallel by the map task. After the map’s processing is finished, the framework sorts the output before feeding it to the reduce task. Because of the presorting feature in the Reduce process, Hadoop results are always orderly within a certain range, giving Hadoop’s MapReduce framework an inherent sorting advantage. The efficiency of MapReduce under the Hadoop framework can be greatly improved by making reasonable use of this property [20]. The commonly used functions and functions of the Map function are shown in Table 1.

3.3. Physical Exercise Data Mining Algorithms. Calculating the distance from the test point to other points from the obtained k data points, assuming that any instance x is described by \( x = \{a_1(x), a_2(x), \ldots, a_n(x)\} \), the distance between the two instances can be expressed as follows:

\[
d = \sqrt{\sum_{i=1}^{n} (a_i(x) - a_j(x))^2}.
\]

(1)

\( D \) represents the distance between two points, and \( a_i(x) \) represents the corresponding value on dimension \( i \) of instance \( X \), which is used to calculate the Euler distance between two instances. The weighted calculation is done by distance, and the calculation formula of the weighted function of distance is as follows:

\[
\hat{f} \leftarrow \text{arg max}_{i} \sum_{i=1}^{k} \omega_i \delta(v, \hat{f}(x_i)).
\]

(2)

Among them, \( \omega_i \) represents the size of the weight calculated according to the distance, and its calculation formula is as follows:

\[
\omega_i = \frac{1}{d(x_q, x_i)^2}.
\]

(3)

\( X_q \) is the prediction point, \( X_i \) is the adjacent point of \( X_q \), and the reciprocal of the distance between the two is calculated as the size of the weight. Of course, other calculation methods can also be used. The weights are weighted according to the mixed Gaussian model, and \( d(x_q, x_i) \) is the Euler distance local model. If \( x_q = x_i \), then \( \hat{f}(x_q) = f(x_i) \), as shown in Figure 6.

In local weighted regression, the weighting of points is calculated according to the distance, and the regression uses the weighted points to calculate the cost. There are many weighting methods for points; the most commonly used is the Gaussian model, such as the following:

\[
h_1 = h(x - x_i) = \exp \left(-k(x - x_i)^2\right).
\]

(4)

In formula (4), \( k \) is a smoothing parameter, determined by the following steps.

Formula (5) is used to calculate the average mathematical expectation of surrounding neighbors.

\[
\mu_x = \frac{\sum h_i x_i}{n}.
\]

(5)
Figure 4: Structure diagram of Hadoop-related components.

Figure 5: MapReduce workflow.

Table 1: Common functions of the Map function.

| Class name              | Function                                                                 |
|-------------------------|--------------------------------------------------------------------------|
| TextInputFormat         | Read a plain text file, the key is the offset of each line, and the value is the content of the line |
| KeyValueTextInputFormat | Read a plain text file, the default first column is key, and the rest is value       |
| SequenceFileInputFormat | Read Hadoop custom datastore binaries                                      |
| NLineInputFormat        | Similar to TextInputFormat, the difference is that the original data can be divided into lines |
| DBInputFormat           | Read data from the database                                              |
Formula (6) is used to calculate the mathematical variance of the neighboring points.

\[ \sigma^2_x = \frac{\sum h_i(x_i - \mu_x)^2}{n}. \]  
(6)

Formula (7) is used to calculate the covariance between two points in the surrounding neighborhood.

\[ \sigma^2_{xy} = \frac{\sum h_i(x_i - \mu_x)(y_i - \mu_y)}{n}. \]  
(7)

The expansion expressions of the formula are as follows:

\[
\begin{align*}
\mu_y &= \frac{\sum h_i y_i}{n}, \\
\sigma^2_y &= \frac{\sum h_i(y_i - \mu_y)^2}{n}, \\
\sigma^2_{yx} &= \sigma^2_y - \frac{\sigma^2_{xy}}{\sigma^2_x}. \\
\end{align*}
\]
(8)  
(9)  
(10)

Using the covariance of the data to calculate and describe the expectation and variance of the data, the expectation is calculated as

\[ \tilde{y} = \mu_y + \frac{\sigma_{yx}}{\sigma^2_x}(x - \mu_x). \]  
(11)

The variance calculation formula is as

\[ \sigma^2_y = \frac{\sigma^2_{yx}}{n^2} \left( \sum i h_i^2 \frac{x - \mu_x}{\sigma^2_x} - \frac{\sum h_i^2(x_i - \mu_x)^2}{\sigma^2_x} \right). \]  
(12)

From the calculated results, the variance-based method can achieve the best results, that is, the value of the minimum extremely smooth coefficient \( k \) around the reference point. Then, the regression function is as follows:

\[ \tilde{f}(x) = \beta_0 + \beta_1 a_1(x) + \cdots + \beta_n a_n(x). \]  
(13)

In formula (13), \( \beta_0 \) is the constant term of the regression, \( \beta_i \) is the regression coefficient, and \( \tilde{f}(x) \) is the predicted value of the regression. The expectations of the first two are shown in the following:

\[ E_1(x_q) = \frac{1}{2} \sum_{x \in k} \left( f(x) - \tilde{f}(x) \right)^2, \]  
(14)

\[ E_2(x_q) = \frac{1}{2} \sum_{x \in D} \left( f(x) - \tilde{f}(x) \right)^2 K(d(x_q, x)). \]  
(15)

The expectation can be obtained as

\[ E_3(x_q) = \frac{1}{2} \sum_{x \in k} \left( f(x) - \tilde{f}(x) \right)^2 K(d(x_q, x)). \]  
(16)

Performing weighting processing on each point, as shown in

\[ \sum_k w_k \left( f(x) - \beta^T x \right)^2 \rightarrow \text{min}. \]  
(17)

The weight of \( y \) can be obtained, as in

\[ w_y = \text{Diag}(w_1, w_2, \cdots, w_n) y. \]  
(18)

Using the weighted points to calculate the regression coefficient \( \beta \), as in

\[ \beta = (WX^T WX^{-1})(WX^T WY). \]  
(19)

4. Physical Exercise Data Mining

The physical fitness test should have become a set of mechanisms to urge college students to participate in physical exercise and a supervising force to strengthen the physical fitness training and training of college students. However, due to various factors, physical education has become an obstacle in the minds of students, and it has developed into a physical exercise that students cannot do in order to pass the exam. In some provinces and cities, “the failure rate of college students’ physical health reaches 10.25%, and some of the students’ functions and quality indicators have declined.” Students do not regard physical activity as a course that they must experience in their life and do not realize the important role of strengthening physical exercise in improving their quality of life. The school also does not realize that the physical fitness test and the development of physical education courses are an important part of the development of sports in colleges and universities nor do they realize that guiding students to the correct physical fitness exercise is a performance of responsibility to the students. Under the current circumstances, the school’s physical education system also needs to be improved in many aspects. Through the continuous reform of physical education and teaching, the development of college sports will be promoted.

4.1. Data Sources. The data content of this experiment includes weight, height, exercise mode, exercise time period, and exercise time length. The data structure of the file is shown in Table 2.

4.2. Data Cleaning. Data cleaning refers to solving the problem of dirty data, eliminating inconsistencies in the data, and at the same time preliminarily formatting the data according to the requirements of data mining, so as to provide a good data source for the final data mining. Therefore, two aspects should be done in data cleaning: (a) removing invalid data and (2) data formatting. Usually, data processing programs are executed based on the database. Database providers such
Table 2: File data structure.

| Number of columns | Name      | Data format |
|-------------------|-----------|-------------|
| 1                 | Height    | ***cm       |
| 2                 | Weight    | **kg        |
| 3                 | Sports    | Running, ball games, others |
| 4                 | Exercise time | Morning (6–8 am), afternoon (5-7 pm), others |
| 5                 | Exercise time | 0–30 min,31–60 min, >60 min |

Table 3: Hive cleaning run test results.

| Number of nodes | First test   | Second test  | Third test  | Average value |
|-----------------|--------------|--------------|-------------|---------------|
| 3               | 2150.38 s    | 2224.78 s    | 2137.54 s   | 2170.9 s      |
| 6               | 1125.86 s    | 1103.47 s    | 1189.13 s   | 1139.5 s      |
| 9               | 754.97 s     | 800.32 s     | 772.51 s    | 775.93 s      |
| 12              | 620.51 s     | 608.92 s     | 578.37 s    | 602.6 s       |

As Oracle and MSSQL provide their own data cleaning tools, such as SISS (SQL Server 2005 Integration Service), which are commonly used data cleaning tools. Most data cleaning tools of this type provide thread-based parallelization, and the efficiency of the overall data cleaning process is directly determined by the hardware system. Since most of the parallelization implementations of the system are based on the parallelization of shared memory, the system scalability is poor. Therefore, in the case of running such data cleaning tools, if want to obtain better performance, it must have a powerful hardware configuration, such as an expensive minicomputer. In addition, since the design of this type of data cleaning tool is based on the realization of relational database at the beginning, it essentially denies the possibility of this type of data cleaning tool to clean unstructured data. Today’s data structure has developed from structured data to multidimensional unstructured data, such as web data. Therefore, such data cleaning tools cannot achieve satisfactory results in the face of massive unstructured data.

In order to set up a data cleaning framework in a Hadoop cluster, the original cleaning process must first be converted into a MapReduce model, and then the Hadoop framework must be used for coding. We decided to use the Hive tool to complete the data cleaning in order to reduce code writing time and simplify the MapReduce code writing process. Hive is a MapReduce query implementation that works like SQL and can use any file in HDFS as a source file for fast MapReduce operations. Because Hive’s main function is to convert SQL-like statements into a series of MapReduce programs, using Hive for data cleaning is a parallelization process based on the Hadoop framework, and the Hadoop framework determines its overall scalability. As a result, using Hive as a data cleaning tool can fully meet the paper’s scalability requirements for data mining.

The purpose of this test is to examine the linear scalability of data cleaning tools based on Hadoop+Hive mode, so 3, 6, and 12 nodes are used for testing.

It can be seen from Table 3 that with the increase of the number of nodes, the test time will gradually decrease, which shows that in the process of data cleaning, nodes can be appropriately added for big data to improve cleaning efficiency.

It can be seen from Figure 7 that when the computing nodes increase to 2 times, 3 times, and 4 times the original number of nodes, respectively, the performance gains obtained are 1.91, 2.80, and 3.60, respectively, and the linear expansion indicators reach 0.955, 0.933, and 0.9. This shows that the data cleaning tool based on Hadoop+Hive has excellent linear scalability. Therefore, in practical applications, the overall mining speed can be improved by adding more computing resources. However, it should be noted that when the computing tasks on each node are already low, the performance increase obtained by adding more computing nodes is often small. Therefore, the expansion capability at this time will no longer be linear, and this state is called the inflection point in this paper. In practical applications, computing resources should be adjusted reasonably according to the operating status of nodes to prevent inflection points due to excess resource increase, resulting in unnecessary cost.

5. Data Mining Algorithm Performance Analysis

In order to test the performance of the physical exercise data mining (PEDM) algorithm, this paper uses the CloudSim simulation platform to simulate the construction of a data center and compares the algorithm with the traditional data mining algorithm. The experimental test includes two aspects: the first type of experiments is mainly compared from the perspective of task completion time, and the second type of experiments is mainly compared from the perspective of equilibrium factors.

It can be seen from the experimental results in Figure 8 that in the early stage of the experiment, the PEDM algorithm even takes more time than the traditional algorithm, but in the later stage of the experiment, the total execution time of the PEDM algorithm is significantly less than that of the traditional algorithm. The main reason is that the
cloning operation in the early stage of the PEDM algorithm leads to the expansion of the feasible solution space, thereby reducing the search speed in the early stage. However, in the later stage of evolution, as the proportion of excellent antibodies in the population gradually expands, the PEDM algorithm can find the optimal solution in a shorter time.

It is not difficult to see from the experimental results in Figure 9 that the curve fluctuation range of the PEDM

Figure 7: Hive linear growth graph.

Figure 8: Comparison of algorithm completion time.
algorithm is the smallest. It shows that the algorithm makes the average load balance of the system better, while the traditional algorithm fluctuates greatly. The maximum value of the traditional algorithm is close to 84%, and the minimum value is about 38%.

As can be seen from Figure 10, when $n = 1000$ and 2000, through the PEDM algorithm, the different computing resources are all above 80%, and they are relatively stable without major fluctuations, and the system load balance can be basically achieved.

6. Conclusions

The purpose of this paper is to look into the data mining algorithm used to analyze students’ physical activity. By investigating the physical fitness and status of current college students, conducting data mining analysis, identifying the factors that influence college students’ physical fitness, and conducting a more detailed analysis of the influencing factors, the algorithm can grasp the current level of physical fitness of college students. The background, significance, and related work of the article are first presented in this article, and it is discovered that current research on students’ physical activity is more concerned with a questionnaire survey of the students themselves, as well as the impact of sports on students. Despite the fact that follow-up experimental research has shown that the algorithm has good performance, it has not been optimized for other parts due to space constraints. In the follow-up research, it is hoped that more research can be done.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors do not have any possible conflicts of interest.

References

[1] C. Helma, T. Cramer, S. Kramer, and L. De Raedt, “Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds,” *Journal of Chemical Information and Computer Sciences*, vol. 35, no. 4, pp. 1402–1411, 2018.

[2] A. Buczak and E. Guven, “A survey of data mining and machine learning methods for cyber security intrusion detection,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016.

[3] L. Xu, C. Jiang, J. Wang, J. Yuan, and Y. Ren, “Information security in big data: privacy and data mining,” *IEEE Access*, vol. 2, no. 2, pp. 1149–1176, 2017.

[4] I. Kavakiotis, O. Tsavel, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, “Machine learning and data mining methods in diabetes research,” *Biotechnology Journal*, vol. 15, pp. 104–116, 2017.
[5] V. Chaurasia and S. Pal, “A novel approach for breast cancer detection using data mining techniques,” Social Science Electronic Publishing, vol. 3297, no. 1, pp. 2320–9801, 2017.

[6] X. S. Yan and L. Zheng, “Fundamental analysis and the cross-section of stock returns: a data-mining approach,” Review of Financial Studies, vol. 30, no. 4, pp. 1382–1423, 2017.

[7] V. Grassmann, M. V. Alves, R. F. Santos-Galduróz, and J. C. F. Galduróz, “Possible cognitive benefits of acute physical exercise in children with ADHD,” Journal of Attention Disorders, vol. 21, no. 5, pp. 367–371, 2017.

[8] V. Grassmann, M. V. Alves, R. F. Santos-Galduróz, and J. C. F. Galduróz, “Possible cognitive benefits of acute physical exercise in children with ADHD,” Journal of Attention Disorders, vol. 21, no. 5, pp. 367–371, 2017.

[9] K. Håkansson, A. Ledreux, K. Daффner et al., “BDNF responses in healthy older persons to 35 minutes of physical exercise, cognitive training, and mindfulness: associations with working memory function,” Journal of Alzheimers Disease, vol. 55, no. 2, pp. 645–657, 2016.

[10] L. C. Melo, J. Dativo-Medeiros, C. E. Menezes-Silva, F. T. Barbosa, C. F. D. Sousa-Rodrigues, and L. A. Rabelo, “Physical exercise on inflammatory markers in type 2 diabetes patients: a systematic review of randomized controlled trials,” Oxidative Medicine and Cellular Longevity, vol. 2017, no. 4, pp. 1–10, 2017.

[11] X.-B. Jin, W.-T. Gong, J.-L. Kong, Y.-T. Bai, and T.-L. Su, “PFVAE: a planar flow-based variational auto-encoder prediction model for time series data,” Mathematics, vol. 10, no. 4, p. 610, 2022.

[12] A. Oliveira de Carvalho, E. Murillo-Rodriguez, N. B. Rocha, M. G. Carta, and S. Machado, “Physical exercise for Parkinson’s disease: clinical and experimental evidence,” Clinical Practice & Epidemiology in Mental Health, vol. 14, no. 1, pp. 89–98, 2018.

[13] R. B. Silva, H. Aldoradin-Cabeza, G. D. Eslick, S. Phu, and G. Duque, “The effect of physical exercise on frail older persons: a systematic review,” The Journal of Frailty & Aging, vol. 6, no. 2, pp. 91–96, 2017.

[14] S. S. Tomazoni, E. Leal-Junior, R. C. Pallotta, S. Teixeira, P. de Almeida, and R. A. B. Lopes-Martins, “Effects of photobiomodulation therapy, pharmacological therapy, and physical exercise as single and/or combined treatment on the inflammatory response induced by experimental osteoarthritis,” Lasers in Medical Science, vol. 32, no. 1, pp. 101–108, 2017.

[15] A. E. Den Heijer, Y. Groen, L. Tucha et al., “Sweat it out? The effects of physical exercise on cognition and behavior in children and adults with ADHD: a systematic literature review,” Journal of Neural Transmission, vol. 124, no. S1, pp. 3–26, 2017.

[16] Y. Kurebayashi and J. Otaki, “Does physical exercise increase brain-derived neurotrophic factor in major depressive disorder? A meta-analysis,” Psychiatry Danubina, vol. 30, no. 2, pp. 129–135, 2018.

[17] V. Mastrorilli, C. Scopa, D. Saraulli et al., “Physical exercise rescues defective neural stem cells and neurogenesis in the adult subventricular zone of Btg1 knockout mice,” Brain Structure & Function, vol. 222, no. 6, pp. 2855–2876, 2017.

[18] E. I. De Bruin, A. R. Formsm, G. Frijstein, and S. M. Börgers, “Mindfull2Work: effects of combined physical exercise, yoga, and mindfulness meditations for stress relieve in employees,” Mindfulness, vol. 8, no. 1, pp. 204–217, 2017.

[19] B. Fenesi, H. Fang, A. Kovacevic, M. Oremus, P. Raina, and J. J. Heisz, “Physical exercise moderates the relationship of apolipoprotein E (APOE) genotype and dementia risk: a population-based study,” Journal of Alzheimer’s Disease, vol. 56, no. 1, pp. 297–303, 2017.

[20] I. B. Henriques, C. E. Keutenedjian Mady, and D. Silvio, “Assessment of thermal comfort conditions during physical exercise by means of exergy analysis,” Energy, vol. 128, pp. 609–617, 2017.