Hindawi
Scientific Programming
Volume 2022, Article ID 3603238, 10 pages
https://doi.org/10.1155/2022/3603238

Research Article

Data Fusion Approach for Managing Clinical Data in an Industrial Environment using IoT

Mrunalini Harish Kulkarni,[1] **Chaitanya Kulkarni**,[2] **K. Suresh Babu**,[3] **Saima Ahmed Rahin**,[4] **Shweta Singh**,[5] **and D. Dinesh Kumar**[6]

[1]*Department of Pharmaceutical Chemistry, School of Pharmacy, Vishwakarma University, Pune, India*
[2]*Computer Engineering, VPKBIET, Baramati, India*
[3]*Department of Biochemistry, Symbiosis Medical College for Women, Symbiosis International (Deemed University), Pune, India*
[4]*United International University, Dhaka, Bangladesh*
[5]*Electronics and Communication Department, IES College of Technology, Bhopal, India*
[6]*Department of Electronics and Instrumentation Engineering, St. Joseph's College of Engineering, OMR Road, Chennai 600119, Tamilnadu, India*

Correspondence should be addressed to Mrunalini Harish Kulkarni; mrunalini.kulkarni@vupune.ac.in and Saima Ahmed Rahin; srahin213012@mscse.uiu.ac.bd

Received 22 March 2022; Revised 18 April 2022; Accepted 30 April 2022; Published 23 May 2022

Academic Editor: Ahmed Farouk

As health issues continue to become more prevalent as the population grows, building a public health network is critical for enhancing the overall health quality of the community. This study offers an Internet of Things (IoT) based health care system that can be employed in the context of community medical care industrial areas. The main focus of this research is to develop a disease prediction strategy that could be applied to community health services using theoretical modelling. Using principal component analysis (PCA) and cluster analysis, an artificial bee colony (ABC) creates a nonlinear support vector machine (SVM) classifier pair. Feature-level fusion analysis was performed to detect probable abnormalities. The results of the experiments reveal that the SVM model offers significant benefits in disease prediction. In the SVM illness prediction model, the ABC algorithm has the best parameter optimization effect in terms of accuracy, time, and other factors. The suggested method outperformed the traditional SVM and BP neural network methods by 17.24 percent and 72.41 percent, respectively. It can lower the RMSE and improve assessment indicators like the precision recall rate and the F-measure, demonstrating the method's validity and accuracy. As a result, it is frequently used in community health management, geriatric community monitoring, and clinical medical therapy in an industrial environment.

1. Introduction

With the rapid development of the global economy and the acceleration of life, increasingly, people have been in a subhealth state for a long time and are prone to chronic diseases. Some fatal diseases are hidden, and the early pathological characteristics are not apparent [1]. Obvious symptoms appear in the late stage, and the best treatment period has been missed at this time. With the acceleration of population ageing, health problems are becoming increasingly prominent, so improving community health infrastructure is very effective practical significance to enhance the overall health quality of the whole people. Based on known published research on the key features of data procurement and management in the IoT in conjunction with data fusion and mining technology, this proposed research aims to investigate support vector machine (SVM) for the projection and diagnosis of public health management to address some underlying complications in this field.

To improve the disease prediction ability of the community, medical and health management system to improve people's lives and health is a very worthwhile research topic. Developments in digital health have permitted the

acquisition of vast volumes of data in clinics, homes, and communities over the last few decades. Activity and metabolic data were collected using wearable sensors. Contextual information has been given by ambient detectors and wearable cameras. Electronic Patient-Reported Outcomes (ePROs) have been collected using smartphones and tablets. Concurrent with this, improvements in machine learning have created an opportunity to extract therapeutically relevant information from vast datasets. The discipline of rehabilitation medicine is changing as a result of these advancements [2]. In recent years, many scholars at home and abroad have carried out in-depth research in this field and achieved rich results. Wallace et al. [3] realized automatic a-trial for fibrillation and coronary heart disease by analyzing the pulse signal. The authors of [4], based on the Internet of Things technology, builds a health management system, uses a variety of machine learning techniques to analyze a variety of disease data sets, and finds that the random forest has a good prediction effect on a variety of diseases, with reasonable accuracy, but the algorithm has poor robustness. Random forests (RF) are multidecision tree ensemble classifiers that train numerous decision trees randomly. The random forest approach is made up of two steps: a training step that creates numerous decision trees and a testing step that categorizes or estimates an outcome variable depending on an input vector. Theoretically, RF is immune to overfitting and is unaffected by noise or anomalies. Furthermore, by lowering generalization defects, it can produce high-accuracy outcomes. However, on the other hand, RF is more inclined to have an elbow point, implying a steeper gradient with more trees. Furthermore, choosing an insignificant explanatory variable increases the likelihood of each tree being more intricate [5].

The Internet of Things (IoT) healthcare system allows for more efficient monitoring and tracking, which aids in better resource management. IoT can effectively track patients from afar as well as provide emergency assistance, which is incredibly beneficial for cardiac patients. The following are some of the primary benefits of IoT in healthcare:

Cost savings: IoT allows for real-time patient tracking, reducing the number of unwanted medical appointments, hospitalizations, and readmissions.

Treatment that is more effective: it allows doctors to make educated decisions supported by facts and ensures greater transparency.

Quicker disease prognosis: using continuous patient monitoring and actual information, doctors can diagnose disease at an earlier phase, well before symptoms appear.

Providing proactive medical care: continuous health monitoring allows for the provision of proactive medical interventions.

Control over drugs and equipment: in the medical industry, managing pharmaceuticals and hospital instruments is a large concern. These are effectively handled and used by connected devices, leading to lesser expenses.

Reduced error: data provided by IoT devices not only aids in a better judgement call but also guarantees that medical operations run smoothly with minimal errors, loss, and system expenses.

Various healthcare sensors generate a plethora of information in healthcare applications. These disparate devices generate information in a wide range of formats. In most clinical decisions, a single source of data might not even be sufficient to reach an appropriate conclusion. These different types of data can be merged for thorough assessment, which aids in the development of a better knowledge of the condition. For both patients and healthcare practitioners, combining information from multiple sources such as clinical databases, sensory equipment, historical, or textual data is critical [6, 7]. Sensor data acquired, merged, and analyzed are critical for diagnosing and treating patients with severe conditions (such as hypertension and diabetes) as well as tracking and assisting the elderly [8]. Dautov et al. suggested a distributed hierarchical data fusion architecture that combines information from multiple sources at every stage of the IoT taxonomy to generate timely and reliable outcomes [9]. Several diverse and complex data sources can be combined, and the measured data can be processed and transported to a superordinated data-science-oriented cloud solution, according to Neubert et al. on a mobile data collection system; their unique concept emphasizes on the incorporation and fusion of several mobile data sources (mDCS) [10]. The authors of [11] established a similar unified management platform, using the back propagation neural network (BP-NN) to analyze the prevalence of chronic diseases (such as heart disease and diabetes) based on the physiological health data of the elderly, with high accuracy to achieve early detection and early treatment of diseases. An input layer, an implicit layer, and an output layer make up a BP neural network. It is a pedagogical approach based on signal transmission, where the signal is carried in two phases: forward and backward propagation, and it is adept at approximating nonlinear mappings with arbitrary precision. If the output value is near the required value, after the reiterating the cycle, the training is completed [12]. Although the functions of the above health management systems are relatively complete, the models involved in data processing and disease prediction are relatively simple. A single type of disease cannot predict potential conditions based on multiple physiological indicators comprehensively. It needs to be improved in practical applications. A single disease cannot comprehensively forecast potential circumstances based on several physiological signs. It requires improvement in practical applications because most diseases follow their own evolutionary principles, which are frequently accompanied by changes in many physiological markers of the human body. To accurately anticipate disease, this necessitates a large volume of patient data. Authors of [13] used machine learning methods (nearest neighbor method, SVM) to predict chronic obstructive disease staging with high accuracy and specific clinical significance. However, this study did not. The theory is integrated into the

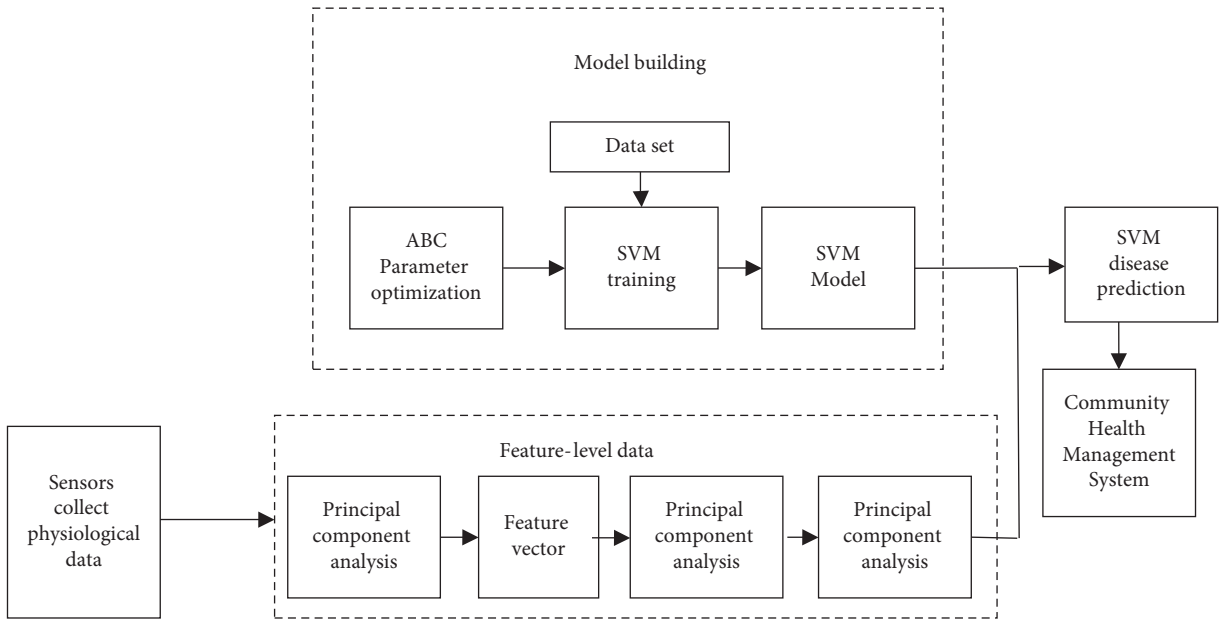

Figure 1: Community disease prediction model.

existing health infrastructure, and the application scenarios are relatively narrow. Author of [14] applied the hybrid machine learning algorithm combined with the Internet of Things technology to predict heart disease, which improved the prediction accuracy to 100%, and the prediction effect was remarkable. Authors of [15] applying machine learning and big data technology to the health care community improve the prediction accuracy. It can be seen that the full use of extensive data analysis technology can analyze the potential disease development law from massive data and then make accurate predictions of the disease [16].

Aiming at some existing problems in this field, based on existing research results, through the research on key technologies of data acquisition [17] and management in the Internet of Things environment, combined with data fusion and mining technology, support vector machine (SVM) is introduced into the disease prediction of community health management. After parameter optimization by the artificial bee colony (ABC) algorithm, the prediction accuracy can be effectively improved. Furthermore, thanks to the self-learning ability of SVM, with the increase of the knowledge base, the disease prediction ability will also be enhanced, which is conducive to promoting the grid construction of community medical care and residents' health management and giving full play to the advantages of community medical care.

The contribution of this paper is in the following features:

(1) An artificial bee colony (ABC) constructs a nonlinear support vector machine (SVM) classifier pair using principal component analysis (PCA) and cluster analysis

(2) To detect potential irregularities, feature-level fusion analysis is used

The rest of the paper follows the following organization: in Section 2, the theoretical aspects of the disease prediction model are being investigated along with parameter optimization; Section 3 is focused on the analysis of the experiments performed, and finally Section 4 concludes the paper.

2. Theoretical Analysis of Disease Prediction

2.1. Community Disease Prediction Model. As shown in Figure 1, the physiological data collected by the sensor extract critical information through the feature-level data fusion method and input the optimized SVM model for disease analysis and prediction, and the final processing results can be fed back to the community health management system.

2.2. Data Feature Extraction. Due to the large amount of data collected by the community IoT terminal, high dimension, and possible unfavorable factors such as accidental errors and noise interference, it is necessary to reduce the size and compress the physiological data before data analysis and remove redundant information and minimize interference to enhance the system efficiency. Principal component analysis (PCA) is a dimensionality reduction algorithm commonly used in data mining [18]. It can use a few linearly independent main components to reflect the most original variables through dimensionality reduction techniques—part of the information [19].

The detection samples with group dimension of nm collected:

$$Y = (y1, y2, y3, \ldots.., yn). \quad (1)$$

In the formula, it is the row and column vector. To uniformly represent different feature dimensions in the sample, it is necessary first to centre the piece.

$$Y^{\sim} = (y_1 - \mu_1, y_2 - \mu_2, \ldots, y_n - \mu_n),$$
$$\mu_i = \frac{1}{m}\sum_{j=1}^{m} y_i^j \left[1, 1, 1, \ldots 1^{U]}\right]; i = 1, 2, ..n. \quad (2)$$

In the formula: μ_i is the mean vector of the physiological data of dimension, and in any size, the sample variance is

$$\mathrm{var}(y_i) = \frac{1}{m}\sum_{i=1}^{m}\left(y_a^{\sim i}\right)^2. \quad (3)$$

When the number of sensors $n \geqslant 2b$, the covariance of dimension and dimension is

$$\mathrm{Cov}(y_a, y_b) = \frac{1}{m}\sum_{i=1}^{m} y_a^{\sim i} y_b^{\sim i}. \quad (4)$$

Let the matrix $D = 1/m\, Y^{\sim U} Y^{\sim}\ C = 1\tilde{m}XT\tilde{}X$ be

$$D = \begin{vmatrix} \frac{1}{m}\sum_{i=1}^{m} y_1^{\sim i} y_1^{\sim i} & \ldots.. & \frac{1}{m}\sum_{i=1}^{m} y_1^{\sim i} y_n^{\sim i} \\ \cdots & \cdots & \cdots \\ \frac{1}{m}\sum_{i=1}^{m} y_n^{\sim i} y_1^{\sim i} & \cdots & \frac{1}{m}\sum_{i=1}^{m} y_n^{\sim i} y_n^{\sim i} \end{vmatrix}. \quad (5)$$

C diagonal elements are the sample variances, and the off-diagonal elements are the covariances. The goal of dimensionality reduction optimization is to reduce the dimensional data. For dimension, the method is to select an ortho-normal unit basis. The pairwise covariance of the data under the linear representation of this set of unit orthonormal basis is the variance with the maximum value. That is, it is a symmetric matrix. The matrix's transition matrix after the basis transformation is also diagonal; such a front row is the selected ortho-normal basis. Since it is a real symmetric matrix, the eigenvectors corresponding to its different Eigen values are orthogonal to each other.

$$Q^U D Q = Q^U \frac{1}{m} Y^{\sim U} Y^{\sim} Q = \frac{1}{m}\left(X^U X\right) = \Delta = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}. \quad (6)$$

In summary, the ortho-normal basis sought is Q, and the data obtained after dimensionality reduction is X.

PCA problem can be attributed to the similar diagonalization problem of real symmetric matrices, convenient for computer implementation. PCA (principal component analysis) is a dimensionality reduction approach applied frequently in data mining. Through dimensionality reduction techniques—part of the information—it can employ some linearly independent critical components to describe the most original variables. This statistical method condenses a group of interrelated variables into a few dimensions that capture a significant portion of the original variables' variability. These dimensions are known as components, and they have the property of gathering strongly correlated variables inside each component while remaining uncorrelated [20]. After PCA dimensionality reduction, the data dimension is reduced, and the redundant horizontal data is removed from the above. The redundant longitudinal information is reflected in the redundant samples. Removing the redundant samples is also a direction that needs to be optimized. The cluster analysis method [21] is used to find the centre of gravity of the sample data, and the representative samples are extracted. The specific steps are as follows.

(1) Randomly select a sample as the cluster centric.

$$h_1, h_2, \ldots, h_j \in Y. \quad (7)$$

(2) For each example, the calculation criterion of the cluster centre $d(i)$ is

$$d^i = \mathrm{argmin}\left\|y^i - h_j\right\|^2. \quad (8)$$

(3) Recalculate the centric

$$h_j = \frac{\sum_{i=1}^{n}\left\{d^i = j\right\} y^i}{\sum_{i=1}^{n}\left\{d^i = j\right\}}. \quad (9)$$

Repeat the above three steps to divide the samples into clusters with different interclass distances L according to the distribution characteristics of the data, then select the pieces closer to the cluster centre according to specific criteria, and discard some redundant sample points on the edge of the cluster. After the data is preprocessed above, the dimensions and the number of samples are reduced, the system overhead and computational complexity are reduced, and the processing efficiency of subsequent work is improved.

2.3. SVM Disease Classification Model. SVM is a collection of related supervised learning methods for classification and regression in diagnosing diseases. SVM maximizes the geometric margin while decreasing the empirical classification error. As a result, SVM stands for maximum margin classifier. SVM is a fundamental approach based on statistical learning theory's ensured risk boundaries, often known as the structural risk minimization principle. The kernel approach allows SVMs to do nonlinear classification effectively by projecting their inputs into feature spaces. The kernel trick enables the classifier to be built sans defining the feature space directly [22]. SVM is a generalized classifier that performs multivariate sensor data classification according to supervised learning, supporting linear and nonlinear types. Its decision boundary is to solve the maximum margin hyperplane for the learning samples [23]. The main idea is to use the maximum margin hyperplane. Correctly classify the physiological data containing different disease information according to the characteristics of the

disease type to achieve the effect of disease prediction [24, 25]. Assuming that there are only two different disease types in the sample space, the sensor feature data can be expressed as

$$U = \{(x1, y1), (x2, y2), \ldots, (xn, yn)\}; xn \in X. \quad (10)$$

In the formula, $yn \in \{+1, -1\}$ is the disease type label. Let z^U be the average vector and c is the intercept, then these two types of samples with different disease information can be obtained from the hyperplane.

$$Z^U x + c. \quad (11)$$

Divide it into two and find the distance from any sample point to the hyperplane:

$$e = \frac{Z^U x + c}{\|Z\|}. \quad (12)$$

Let the point above the hyperplane correspond to the disease label as $y_n = 1$, and the point below corresponds to $y_n = 1$. Define the function interval as

$$\hat{s} = \frac{y_i Z^U x + c}{\|Z\|} = \frac{s}{\|Z\|}; i = 1, 2, \ldots, n. \quad (13)$$

SVM optimization problem: it is necessary to find the hyperplane determined by Z, c so that the distance between the nearest support vector and this hyper plane is as considerable as possible, that is,

$$\min_{Z,c}; \text{s.t}\, \hat{s} \} = \min_{Z,c}\left(\frac{s}{\|Z\|}\right),$$
$$\text{s.t}\, y_i (Z_0 x_i + c_o) \geq s; i = 1, 2, \ldots, n. \quad (14)$$

To maximize the interval, it is only necessary to maximize $\|Z\|^{-1}$, which is equivalent to

To minimize $\|W\|^{-1}$, let $Z = Z_0 s$; $c = c_0$, (14) is equivalent to

$$\min_{Z,c}\left(\frac{1}{2}\|Z\|^2\right) \text{s.t}\, y_i (Z_0 x_i + c_o) \geq 1; i = 1, 2, \ldots, n. \quad (15)$$

According to the transformation of the Lagrange multiplier method, the dual problem of equation (15) is obtained. After solving αi, we can get the optimal hyperplane

$$f(x) = z_0^U x + c_0 = \sum_{i=1}^{m} \alpha_i y_i x_i^U x + c_0. \quad (16)$$

Equations (15) and (16) are established on the premise that the KKT (Karush–Kuhn–Tucker) condition is satisfied.

The above is the linear solution process of the SVM disease prediction model. However, in practice, the physiological data collected by multisensors are generally nonlinearly distributed, and there may be multiple disease types in the sample. For such nonlinear models, the kernel function (kernel function) maps the sample space to the kernel space for solving [24], that is, replace the part in equation (16) with $l(x_i; x_j)$.

2.4. Parameter Optimization of Disease Prediction Models. The key to affecting the prediction effect of the SVM classifier is to select an appropriate kernel function model and its corresponding parameters according to the characteristics of the physiological data. The Gaussian kernel function (RBF) is popular in machine learning, and its performance mainly depends on c,g parameter [26]. Among them is the penalty coefficient. If the value of this parameter is too high, overfitting may quickly occur; otherwise, underfitting will quickly occur, resulting in poor data generalization ability; the RBF form is

$$l_{RBF}(x_i, x_j) = \exp\left|-\frac{e(x_i, x_j)^2}{2\sigma^2}\right| = \exp\left[-\Upsilon c(x_i, x_j)^2\right]. \quad (17)$$

The h parameter determines the distribution of the original data mapped to the new feature space, which is negatively correlated with the number of support vectors, which directly affects the algorithm's speed. To find the optimal parameter d, h, you can use the ABC algorithm for parameter optimization [27]. This algorithm is a parameter optimization algorithm proposed by Karaboga in 2005 to solve the multivariable parameter optimization problem [28]. It has been widely used in many fields, including image processing and numerical optimization. [26, 29], and the algorithm performs better than other heuristic algorithms such as particle swarm optimization (PSO) algorithm and genetic algorithm (GA) in multidimensional data processing [26] ABCSVM is the disease prediction process. Ant colony optimization (ACO) is a probability-based optimization approach that is intended to solve computational problems and discover the ideal route using graphs. ACO is capable of working more proficiently and precisely than GA, attributed to the reason that determining the optimal path involves less calculation time and iterations. Furthermore, the accuracy of ACO is demonstrated by the optimal path discovered in each time run [30]. ACO has the ability to cluster and construct routes, and PSO is simple to implement. However, due to its poor exploration, PSO has issues with parameter selection [31]. First, according to the characteristics of physiological data d; the highest sum combination ensures that the model works with the optimized parameters to improve the accuracy of disease classification prediction.

3. Experimental Analyses

To verify the validity and accuracy of the method in this paper, the experimental data were used the physical examination monitoring data of 5 061 cases in a community in Shanghai in 2017, covering residents of all ages from 16 to 90 years old. Some experimental data are shown in Table 1. For the convenience of discussion, in the experiment, four common diseases, including renal function impairment, dyslipidemia, metabolic syndrome, and diabetic nephropathy, were selected as the research objects, and the validity of the diagnostic prediction model was verified.

Table 1: Physiological datasets used in the experiment (partial samples).

SN	G	AGE	CREA	UA	APOA	APOB	GLU	LDL	UREA	CH
1	F	65	66	261	1.34	0.79	4.9	2.92	5.7	4.81
2	M	61	45	282	1.25	0.76	4.7	3.18	5.4	4.86
3	F	59	63	419	1.37	0.97	4.6	3.04	4.9	3.92
4	M	63	60	274	1.46	1.04	10.7	2.21	4.8	4.79
5	F	76	89	294	1.38	0.9	4.7	1.05	6.5	4.3
6	M	65	69	401	1.57	0.66	4.7	2.47	5.8	4.79
...	...	...	...	...	...	...	...	...	...	...

3.1. Sample Encoding. The sample coding requirements can conveniently represent the type of disease and reasonably determine the kind of disease $24 = 16$ and its severity in the sample according to the coding value. Considering that there may be many different diseases in the same model, to facilitate the computer for processing, we first use 4-bit binary code for the disease type (B1~B4, respectively, represent diabetic nephropathy, metabolic syndrome, dyslipidemia, and renal function damage). For standard samples, the code value is set to 0×0. The above code can indicate the possibility of disease (i.e., hexadecimal 0×0~0×F), and the sample codes of 5 representative disease combinations are shown in Table 2.

After removing outliers from the original data (missing critical data is considered outliers), we use the *t*-SNE dimensionality reduction toolbox to map the 8-dimensional raw data to the 2-dimensional space visualization. In the figure, there are a total of 10 types, colors represent ten different disease combinations *D*, respectively, blue represents a standard sample (i.e., the code value is 0×0), and red represents that the model suffers from the four diseases in Table 2 at the same time (i.e., the coded value is 0×0).

The coding value is 0×F. The gradient process of the color bar from blue to red from top to bottom corresponds to the gradual increase in the types of disease suffered by the sample. It can also be intuitively understood that the warmer the color, the worse the health status of the model.

3.2. Data Feature Extraction. Take the 8-dimensional data of CRE~CH in Table 1 and standardize it according to formula (2) and then carry out PCA processing according to procedures (3)~(6) to obtain the characteristic principal component N of the physiological index and its contribution rate as shown in Figure 2 with Table 3. The curve represents the cumulative contribution rate of the first k main components $R = \sum_{k=1}^{N} r_k\,(k = 1, 2, 3, \ldots, n)$ is the contribution rate of the first principal component. Current principal components when the cumulative contribution rate of the parts is greater than the set value, it can be considered that the information contained in the first k main components is sufficient to represent the entire set of data. The calculation shows that the contribution rate of the first six principal components reaches 97.2%. Central component 6 can contain most of the information about the sample.

After PCA processing, the data dimension is reduced to 6 sizes, and then crucial samples are screened out through cluster analysis, further improving the ant interference ability. The cluster analysis dendrogram of sample *S* is shown in Figure 3 with Table 4. To express the distribution law of sample data more intuitively, only 30 leaf nodes are shown here (one leaf node may correspond to multiple sample points). In the dendrogram, it can be observed that, according to the sample, the data distribution features can be subdivided into 5 clusters in total.

The clustering results are visualized by t-SNE dimensionality reduction. Different symbols in the figure represent

Table 2: Sample codes for 5 representative disease combinations.

Sample encoding	Disease type			
	B4	B3	B2	B1
0*0	N	N	N	N
0*1	N	N	N	Y
0*2	N	N	Y	Y
0*4	N	Y	N	N
0*8	Y	N	N	N

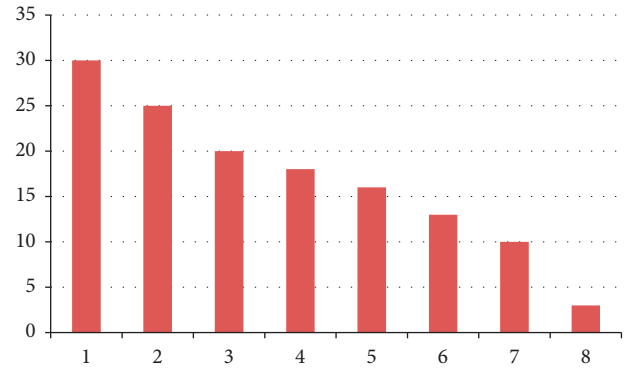

Figure 2: Principal components and contribution rates in physiological indicators.

Table 3: Principal components and their contribution rates in physiological indicators.

Serial	*R*
1	30
2	25
3	20
4	18
5	16
6	13
7	10
8	3

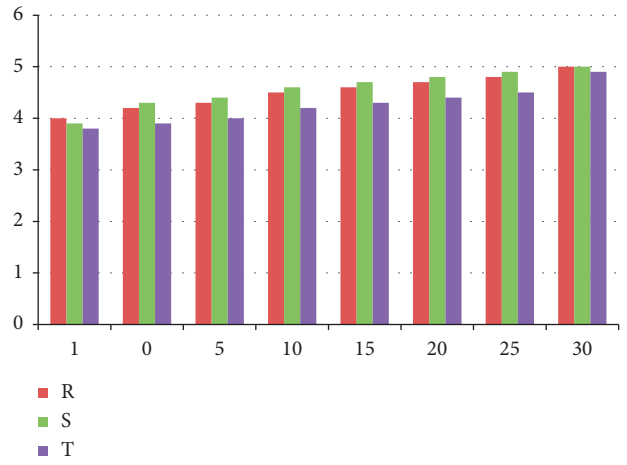

Figure 3: Cluster analysis dendrogram for samples.

Table 4: Cluster analysis dendrogram for samples.

Serial	R	S	T
1	4	3.9	3.8
0	4.2	4.3	3.9
5	4.3	4.4	4
10	4.5	4.6	4.2
15	4.6	4.7	4.3
20	4.7	4.8	4.4
25	4.8	4.9	4.5
30	5	5	4.9

clusters with other characteristics. The distance from the cluster centre d_i ($i = 1, 2,\ldots,5$) is the cluster radius of s_i ($i = 1, 2,\ldots,5$), and the sample points located within the radius are selected as the follow-up experimental data and will not be in the cluster. Points within the class radius are regarded as redundant and discarded, and 1 598 sets of physiological data are finally selected.

In the above five types of samples, 98% of the data of every kind of sample are randomly selected as the training sample $S1$, and the remaining 2% of the data are used as the test sample $S2$ to test the model, and finally, $S1 = 1\,569$ group and $S2 = 29$ groups.

3.3. Analysis of Disease Diagnosis Results. To more intuitively reflect the superiority of SVM in disease analysis, the BP-NN prediction is used as a comparison, and the results are shown in Figure 4 and Table 5. In Figure 4, the three curves represent the prediction effect of BP-NN and the prediction of the SVM model. It can be seen from the comparison that the prediction effect of the SVM model is the best, and the accuracy rate can reach 75 : 86% (22 = 29). In contrast, the disease prediction effect of BP-NN is poor, and the accuracy ratio is only 20 : 69% (6 = 29). Figure 5 and Table 6 shows the root mean square error (RMSE) of the top n predicted values for the two forecasting algorithms.

$$e_{\mathrm{RMSE}} = \left(\frac{1}{2}\left(\hat{x} - x\right)^2\right)^{1/2}. \tag{18}$$

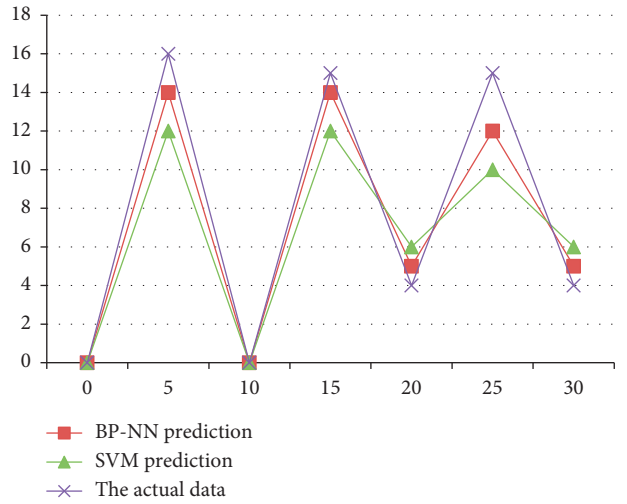

Figure 4: Comparison of the prediction effects.

Table 5: Comparison of the prediction effects of the three diagnostic models.

Serial	BP-NN prediction	SVM prediction	The actual data
0	0	0	0
5	14	12	16
10	0	0	0
15	14	12	15
20	5	6	4
25	12	10	15
30	5	6	4

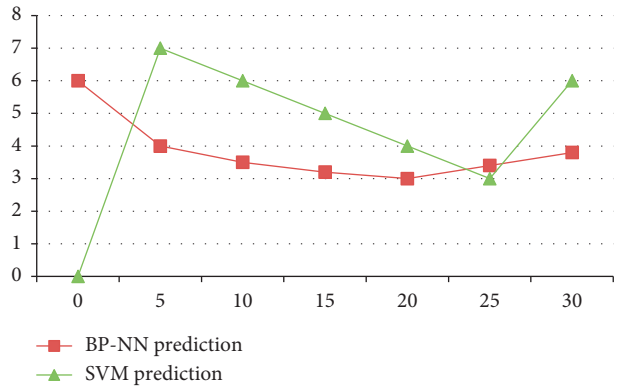

Figure 5: RMSE curves.

Here, $\hat{x}$ and x are the predicted and actual values, respectively. It can be seen from Figure 6 that, in the previous sample, the eRMSE of the SVM model is 0 because the disease type code output by the model in the subsequent false detection is different from the actual value the significant difference in disease type coding leads to a higher RMSE. Still, the performance of this situation in the latter samples is not much different from that of BPNN, which also reflects the importance of SVM parameter optimization from the side. Due to the fuzzy mapping of BP-NN, however, SVM can output discrete and definite value ranges and can accurately output specific sample codes.

Table 6: RMSE curves.

Serial	BP-NN prediction	SVM prediction
0	6	0
5	4	7
10	3.5	6
15	3.2	5
20	3	4
25	3.4	3
30	3.8	6

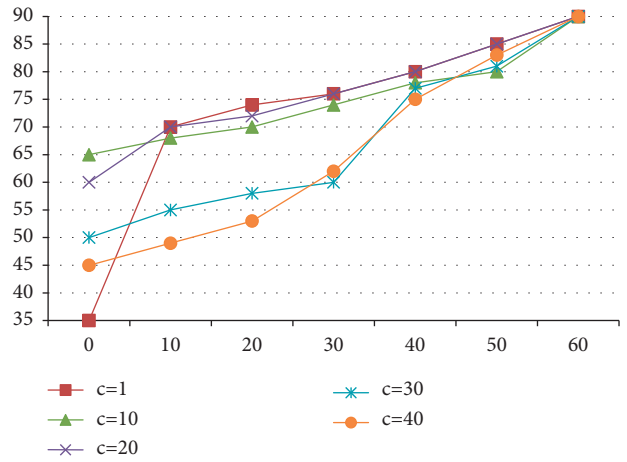

Figure 6: The influence of c,g parameters on the prediction accuracy of ABC-SVM.

Table 7: Accuracy of c,g parameters of ABC-SVM.

Serial	$c=1$	$c=10$	$c=20$	$c=30$	$c=40$
0	35	65	60	50	45
10	70	68	70	55	49
20	74	70	72	58	53
30	76	74	76	60	62
40	80	78	80	77	75
50	85	80	85	81	83
60	90	90	90	90	90

The influence of different c,g parameters on the SVM disease prediction model is shown in Figure 6 and Table 7. It can be seen from the figure that, for the same g parameter, the accuracy rate η of the prediction result is almost unrelated to the value of the c parameter

From the perspective of the entire interval, for the same parameter, the accuracy rate tends to increase with the increase of the parameter, and there may be multiple extreme points in the curve, so how to find the globally optimal powerful moment is the key to model optimization.

The prediction effect of the SVM disease prediction model after optimization by different optimization algorithms is shown in Figure 7 and Table 8. It can be seen from the figure that the disease prediction effect of the PSO algorithm optimization model (PSO optimization, PSO-OPT) is the worst. The total accuracy rate is only 13 : 79%, the GA algorithm optimization model (GA optimization, GA-OPT) has a slightly better effect, the accuracy rate is 79 : 31%, the

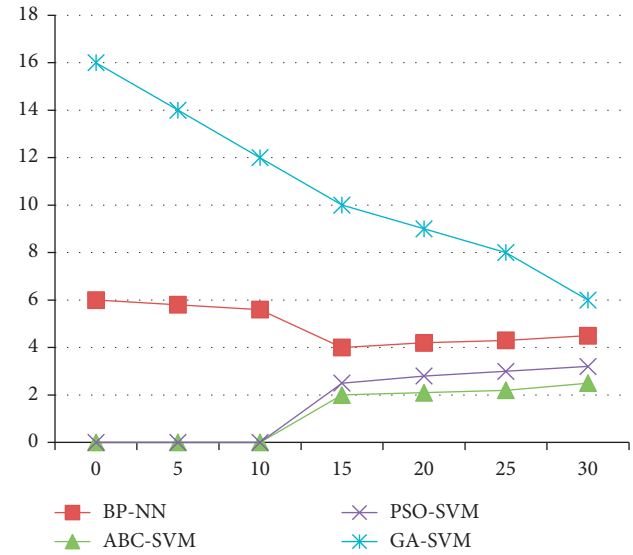

Figure 7: Comparison of the optimization effects of the three optimization algorithms.

Table 8: Comparison of the optimization effects of the three optimization algorithms.

Serial	BP-NN	ABC-SVM	PSO-SVM	GA-SVM
0	6	0	0	16
5	5.8	0	0	14
10	5.6	0	0	12
15	4	2	2.5	10
20	4.2	2.1	2.8	9
25	4.3	2.2	3	8
30	4.5	2.5	3.2	6

ABC algorithm (ABC optimization, ABC-OPT) has the best optimization effect, the accuracy rate can accurately predict the disease types of 93 : 10% of the samples, with the shortest time t and the highest efficiency. The final results are shown in Table 9.

$$Q = \mathrm{Dr} = (\mathrm{Dr} + De),$$

$$S = \frac{Dr}{Dt}, \tag{19}$$

$$G = \frac{2QS}{(S+Q)}.$$

The RMSE comparison of different disease prediction methods is shown in Figure 7. It can be seen from the figure that the optimized SVM model has the lowest eRMSE and apparent advantages.

The prediction effects of different optimization algorithms for different disease combinations are compared in terms of the precision P of disease prediction, the recall rate R, and the F metric $F3$.

Let DP denote the set of disease prediction results in the test sample, DR indicates the actual disease condition in the test sample, Dr is the number of sample codes in DP\DR be

Table 9: Parameters for optimization algorithms.

Optimization algorithm	c	g	t/s	η
GA-OPT	89 : 72	12 : 36	555 : 70	79 : 31% (23 = 29)
PSO-OPT	40 : 69	0.720138889	872 : 06 : 00	13 : 79% (4 = 29)
ABC-OPT	89 : 76	91 : 77	138 : 10 : 00	93 : 10% (27 = 29)

the number of elements in DR and be the number of elements in the set A (and) number of sample codes.

This paper compares different forms of algorithms for data fusion in clinical data, out of which ABC-SVM shows better results compared to the traditional SVM model and BP-NN method. It can lower the RMSE and improve assessment indicators like the precision recall rate and the F-measure, demonstrating the method's precision and credibility.

4. Conclusions

As the population ages, health issues become more prevalent; hence, enhancing community health infrastructure is critical for improving the overall health quality of the populace. The occurrence and development of most diseases follow their unique evolutionary laws, which are often accompanied by changes in various physiological indicators of the human body. Through theoretical modelling, a disease prediction method that can be applied to community health services is proposed. Starting from the health index data of daily life, it can automatically analyze potential health problems, conducive to the timely understanding of human health status and taking corresponding measures. Support vector machine (support vector machine, SVM) is incorporated into the disease prediction of public health management, aiming at some current challenges in this field, based on established scientific data, through research on important features of data acquisition and management in the Internet of Things environment, combined with data fusion and mining technology. The prediction accuracy can be effectively increased after optimizing the process parameters using the artificial bee colony (ABC) technique. Furthermore, because of SVM's self-learning capacity, disease prediction ability will improve as the experience and knowledge grows, which is advantageous in encouraging grid construction of community healthcare and residents' health management, as well as fully exploiting the benefits of public medical care. The experimental results show that the SVM model has outstanding advantages in disease prediction. The ABC algorithm has the ideal parameter optimization effect in the SVM disease prediction model inaccuracy, time-consuming, etc. The recognition accuracy rate reaches 93.10%, 17.24%, and 72.41% higher than the traditional SVM model and BP-NN method, respectively. It can reduce the RMSE and improve the evaluation indicators such as precision, recall rate, and F-measure, which fully proves the method validity and accuracy. The method can be described by a rigorous mathematical model, with fast execution speed, high efficiency, and accessible computer implementation. It has broad application prospects and promotion in improving people's quality of life and health and disease prevention.

4.1. Future Scope. This study has a wide range of potential applications and promotional values in addition to enhancing people's quality of life, health, and illness prevention. However, limited by the current level of economic development, sensor objective factors such as technology, new materials, and their preparation processes, coupled with the variety of diseases and complex pathogenic mechanisms in practical applications, make it still challenging to popularize, and use this method. The planned research will focus on improving sensor technology, reducing production costs and making some medical testing instruments truly wearable and affordable, and improving the prediction accuracy and generalization ability of disease prediction models.

Data Availability

The data shall be made available upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] W. Huang, "Research on user satisfaction of older community care based on structure equation," in *Proceedings of the 2012 Fourth International Symposium on Information Science and Engineering*, pp. 489–492, Washington, DC, USA, December 2012.

[2] P. Bonato, "Keynote: digital health technologies and their role in the development of precision rehabilitation interventions," in *Proceedings of the 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*, p. 200, Kassel, Germany, March 2021.

[3] R. B. Wallace, F. Horsfall, R. Goubran, A. El-Haraki, and F. Knoefel, "The challenges of connecting smart home health sensors to cloud analytics," in *Proceedings of the 2019 IEEE Sensors Applications Symposium (SAS)*, pp. 1–5, Sophia Antipolis, France, March 2019.

[4] A. Jagatheesan, S. Maragathavel, M. Sivapurapu, and J. Lee, "Drops: a multi-producer and multi-consumer data sharing framework with human experience," in *Proceedings of the 2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC*, pp. 601-602, CCNC), Las Vegas, NV, USA, January 2015.

[5] H. Byeon, "Is the random forest algorithm suitable for predicting Parkinson's disease with mild cognitive impairment out of Parkinson's disease with normal cognition?"

International Journal of Environmental Research and Public Health, vol. 17, no. 7, p. 2594, 2020.

[6] M. G. Seneviratne, M. G. Kahn, and T. Hernandez-Boussard, "Merging heterogeneous clinical data to enable knowledge discovery," in *Proceedings of the BIOCOMPUTING 2019: Proceedings of the Pacific Symposium*, pp. 439–443, 2019.

[7] S. Kalamkar, "Clinical data fusion and machine learning techniques for smart healthcare," in *Proceedings of the 2020 International Conference on Industry 4.0 Technology (I4Tech)*, pp. 211–216, IEEE, Pune, India, February 2020.

[8] S. Vitabile, M. Marks, D. Stojanovic et al., "Medical data processing and analysis for remote health and activities monitoring," in *High-performance Modelling and Simulation for Big Data Applications*, pp. 186–220, Springer, Cham, Switzerland, 2019.

[9] R. Dautov, S. Distefano, and R. Buyya, "Hierarchical data fusion for smart healthcare," *Journal of Big Data*, vol. 6, no. 1, pp. 19–23, 2019.

[10] S. Neubert, A. Geißler, T. Roddelkopf et al., "Multi-sensor-fusion approach for a data-science-oriented preventive health management system: concept and development of a decentralized data collection approach for heterogeneous data sources," *International Journal of Telemedicine and Applications*, vol. 2019, Article ID 9864246, 18 pages, 2019.

[11] R. Gargees, J. Keller, and M. Popescu, "Early illness recognition in older adults using transfer learning," in *Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1012–1016, Kansas City, MO, USA, November 2017.

[12] N. Jiang, Z. Zhao, and P. Xu, "Predictive analysis and evaluation model of chronic liver disease based on BP neural network with improved ant colony algorithm," *Journal of Healthcare Engineering*, vol. 2021, Article ID 3927551, 7 pages, 2021.

[13] P. Gope, Y. Gheraibia, S. Kabir, and B. Sikdar, "A secure IoT-based modern healthcare system with fault-tolerant decision making process," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 3, pp. 862–873, 2021.

[14] Q. Chen, W. Wang, F. Wu et al., "A survey on an emerging area: deep learning for smart city data," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 3, no. 5, pp. 392–410, Oct. 2019.

[15] M. Song, Z. Yang, A. Baird et al., "Audiovisual analysis for recognising frustration during game-play: introducing the multimodal game frustration database," in *Proceedings of the 2019 8th International Conference on Affective Computing and Intelligent Interaction*, pp. 517–523, ACII), Cambridge, UK, September 2019.

[16] R. Williams and J. Leonard, "Surgery benefits from defense technology: dual use applications of wright laboratory avionics technologies to computer assisted minimally invasive surgery (CAMIS)," *IEEE Aerospace and Electronic Systems Magazine*, vol. 9, no. 10, pp. 3–6, 1994.

[17] M. Fahim and A. Sillitti, "Anomaly detection, analysis and prediction techniques in IoT environment: a systematic literature review," *IEEE Access*, vol. 7, pp. 81664–81681, 2019.

[18] L. Sevrin, N. Noury, N. Abouchi, F. Jumel, B. Massot, and J. Saraydaryan, "Characterization of a multi-user indoor positioning system based on low cost depth vision (Kinect) for monitoring human activity in a smart home," in *Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5003–5007, EMBC), Milan, Italy, August 2015.

[19] M. Rosenberg and M. A. Williams, "10 bringing global health home," in *Howard Hiatt: How This Extraordinary Mentor Transformed Health with Science and Compassion*, pp. 153–164, MIT Press, Cambridge, MA, USA, 2018.

[20] Z. Zhang and A. Castelló, "Principal components analysis in clinical studies," *Annals of Translational Medicine*, vol. 5, no. 17, p. 351, 2017.

[21] I. Khlobystov and L. Horoshkova, "7 inequality in the distribution of income as at threat to the sustainable development of united territorial communities in Ukraine," in *Economic Inequality–Trends, Traps and Trade-Offs*, pp. 123–138, River Publishers, Denmark, 2021.

[22] V. A. Kumari and R. Chitra, "Classification of diabetes disease using support vector machine," *International Journal of Engineering Research in Africa*, vol. 3, no. 2, pp. 1797–1801, 2013.

[23] L. Lawrence, *Weed; Lincoln Weed, Ending Medicine's Chronic Dysfunction: Tools and Standards for Medical Decision Making*, Morgan & Claypool, San Rafael, CA, USA, 2021.

[24] A. Gupta and L. K. Awasthi, "P4P: ensuring fault-tolerance for cycle-stealing P2P applications," in *Proceedings of the 2007 International Conference on Grid Computing & Applications, GCA 2007*, pp. 151–158, Las Vegas, NV, USA, June 2007.

[25] A. Mehbodniya, J. L. Webber, M. Shabaz, H. Mohafez, and K. Yadav, "Machine learning technique to detect sybil attack on IoT based sensor network," *IETE Journal of Research*, pp. 1–9, 2021.

[26] A. Gupta and L. K. Awasthi, "Peer enterprises: possibilities, challenges and some ideas towards their realization," in *On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops*, pp. 1011–1020, Springer, Berlin Germany, 2007.

[27] S. Deshmukh, K. Thirupathi Rao, and M. Shabaz, "Collaborative learning based straggler prevention in large-scale distributed computing framework," in *Security and Communication Networks*, M. Kaur, Ed., vol. 2021, Article ID 8340925, 9 pages, 2021.

[28] B. Gulzar and A. Gupta, "DAM: a theoretical framework for SensorSecurity in IoT applications," *International Journal of Next-Generation Computing*, vol. 12, no. 3, 2021.

[29] M. P. Bhandari and S. Hanna, "12 social responsibility as a tool for the human resources policy development and reducing inequalities on tourism industry," in *Inequality–the Unbeatable Challenge*, pp. 293–306, River Publishers, Denmark, 2021.

[30] N. B. Sariff and N. Buniyamin, *Genetic Algorithm Versus Ant Colony Optimization Algorithm*, Scitepress, Setúbal, Portugal.

[31] S. Samsuddin, M. S. Othman, and L. M. Yusuf, "A review of single and population-based metaheuristic algorithms solving multi depot vehicle routing problem," *International Journal of Software Engineering and Computer Systems*, vol. 4, no. 2, pp. 80–93, 2018.