Reinforcement Learning Improves SVM-Driven Algorithms for Classifying Multi-Sensor Data for Medical Monitoring

Zhiwei Xuan, Yajie Liu*
School of Information Engineering, Henan Vocational University of Science and Technology, Zhoukou, China

***Abstract*—Multi-sensor data in medical monitoring includes waveform changes in physiological signals and time-series characteristics of disease progression. These features typically exhibit high-dimensionality, large-scale, and time-varying characteristics. Nonlinear relationships exist between these features, increasing the difficulty of data processing and feature extraction, thereby reducing the classification capabilities of related algorithms. This study proposes a multi-sensor data classification processing method in medical monitoring based on reinforcement learning improved SVM. The algorithm employs the DBSCAN algorithm combined with Euclidean distance for clustering and data collection of multi-sensor data. Discrete wavelet transform is used to remove interference noise from the data, followed by convolutional neural networks for signal feature extraction from the denoised data. The Q-learning algorithm in reinforcement learning is used to improve the traditional SVM, with the extracted signal features input into the improved SVM. The classification results of medical monitoring multi-sensor data are output via a regression function. The experimental results show that the denoising results of medical monitoring data of the method are high, the signal-to-noise ratio is high, and the Kappa coefficient reaches up to 0.98. Therefore, it shows that the method can accurately classify medical monitoring multi-sensor data.**

Keywords*—Reinforcement learning; improved SVM; medical monitoring; multi-sensor; data; classification processing

I. INTRODUCTION

Medical monitoring systems [1] are increasingly being used to monitor patient health data. These include electrocardiograms (ECG), electroencephalograms (EEG), blood pressure, blood oxygen saturation, body temperature, and a variety of other physiological signals [2]. Through the processing of monitoring data, statistics on patient cases and health status [3] can be achieved to help doctors make more accurate clinical judgments and decisions; the abnormal classification of medical signals can also promote the development of medical intelligence, thus promoting the progress and innovation of medical technology [4]. The continuous improvement and application of medical signal processing technology are expected to bring more possibilities for clinical medicine and promote the development of the healthcare industry toward intelligence, personalization, and precision. Accurate classification of medical data can help improve the automation level of the medical monitoring system, reduce the workload of healthcare workers, and improve the efficiency and quality of medical services [5]. In addition, the relevant methods can also provide more objective and accurate data support for medical research and promote the progress of medical research and the innovation of clinical practice.

With the continuous development of medical monitoring, multi-sensor data analysis has become an important part of medical diagnosis. Traditional multi-sensor data classifications for medical monitoring often have problems such as low classification accuracy and slow processing speed [6]. Support vector machines (SVMs) are widely used as an effective classification algorithm in medical data classification. However, traditional SVM methods still face many challenges when processing complex and variable data, such as noise interference and inaccurate feature extraction.

Therefore, the research on multi-sensor data classification and processing methods for medical monitoring is not only of great significance for improving the function and performance of medical monitoring systems, but also provides technical support and theoretical guidance for improving medical services, promoting medical research, and safeguarding patients' health [7]. Relevant experts have been continuously exploring the field of data classification.

Singh and Khaiyum [8] proposed a data stream classification method based on the concept of drift, introducing an incremental semi-supervised learning model to regularize neural networks by incorporating auxiliary information such as label merging or pairwise constraints. However, due to the algorithm's efficiency and sensitivity to parameters, it is challenging to achieve better results in highly nonlinear environments with multimedia sensor data. Kenger and Ozceylan [9] proposed a data classification method based on mathematical modeling and improved online learning, using I0L_GFMM to generate initial cluster centers for the MILP model to enhance its efficiency, and combining fuzzy minimum and maximum neural networks with the MILP model. The proposed hybrid model demonstrated its applicability on both real and synthetic datasets. However, the aforementioned methods do not account for noise interference in medical data and are not suitable for high-noise environments in medical monitoring data. Zhai et al. [10] proposed a data classification method based on generative model diversity oversampling. When classifiers are faced with imbalanced multi-sensor-collected medical monitoring datasets, they typically favor the majority category, leading to poor classification performance. Liang et al. [11] proposed a data classification method based on flow regularization to overcome data quality issues and the high labeling cost of extracting data sample labels. However, while

This work was sponsored by Henan Provincial Higher Education Key Research Project Program (Project number of the fund：24B510006).

this classification algorithm can handle specific types of label noise, identifying the type of noise present in each medical data point remains a challenge.

Aiming at the problems of existing methods, this study proposes an improved SVM based on reinforcement learning for the classification of multi-sensor monitoring data. Compared to existing research, the reinforcement learning-enhanced SVM method proposed in this study aims to address the issue of insufficient classification accuracy in traditional algorithms under feature selection and imbalanced dataset scenarios. This is achieved through DWT denoising, CNN feature extraction, and Q-learning optimization. By combining the Discrete Wavelet Transform (DWT) with Multiresolution Analysis (MRA), precise denoising of medical monitoring data is realized, thereby improving the signal-to-noise ratio of the data. Furthermore, this method leverages the advantages of reinforcement learning to enhance the classification performance of medical data through interaction between the agent and the environment. By incorporating the Q-learning algorithm to improve the SVM, the method continuously learns from interactions with the environment to identify the optimal classification strategy, addressing the limitations of traditional support vector machines when processing multi-sensor data. The introduction of reinforcement learning algorithms enables decision automation in exploring optimal feature selection, optimizing model parameters, and balancing classification boundaries, thereby enhancing the model's performance and adaptability.

The subsequent sections of this study are arranged as follows: Section II introduces multi-sensor data acquisition and clustering methods. Section III elaborates on the data denoising process. Section IV discusses CNN feature extraction techniques. Section V provides a detailed explanation of the SVM classification model improved by Q-learning. Section VI validates the effectiveness of the method through experiments. Finally, Section VII summarize the research results and looks forward to future directions.

II. Multi-Sensor Data Acquisition and Clustering in Medical Monitoring

Multi-sensor data involved in medical monitoring are usually characterized by high dimensionality, large scale, and time-varying nature, such as waveform changes in physiological signals and time-series features of disease progression. The high dimensionality and complexity of such data lead to a reduced ability to categorize the data. The use of multi-sensors to collect data can obtain more dimensional information, which helps to improve the accuracy and stability of classification algorithms and enhance the effectiveness and reliability of classification. Therefore, in this study, multi-sensor data acquisition in medical monitoring is used to ensure data integrity and to support the accuracy of feature extraction through sufficiently acquired data, which can help to assess the patient's health status and process it for exhaustive classification more accurately.

The DBSCAN algorithm can discover the existence of potential clusters based on the density between data points, which helps to reveal hidden patterns and structures in the data. In medical monitoring, different types of medical surveillance sensing data may correspond to different health states or disease characteristics, which can be of help to better understand, by discovering clusters in the data. The DBSCAN algorithm identifies whether a data point is a core object or a boundary point, and from this, clusters with different levels of density are constructed. By identifying the core objects, the clusters to which the data points belong can be better distinguished, thus facilitating subsequent data classification tasks.

The DBSCAN algorithm divides the medical monitoring multi-sensor data into classes; those isolated few signal points will be considered as outliers to be eliminated, and the remaining data clusters as useful data. For the data object p, the ε neighboring object is denoted as $N_\varepsilon(p)$ and $N_\varepsilon(p)=\{q\in D \mid dist(p,q)\le \varepsilon\}$. For a given value $MinPts$, the object q is said to be directly density reachable from q if it is a neighboring object of the object ε of p. The point p is said to be directly density reachable from q.

For a given ε and $MinPts$, if there exists an object O in the set of data objects D such that the objects p and q are density reachable from O concerning the densities ε and $MinPts$,then the objects p and q are said to be density-connected about ε and $MinPts$.

For the set of data objects D, the cluster C is a non-empty subset of the set D that satisfies the following conditions:

1) $\forall p,q$, if $p\in C$, p and q are density reachable, then $q\in C$.

2) $\forall p,q\in C$, p and q are density-connected.

The DBSCAN algorithm requires two important parameters, a point p neighborhood ε radius, and the minimum number of points contained in the neighborhood $MinPts$. Firstly, the data set of a medical sensor monitoring object is specified, from which any data object p is selected, and all the objects are scanned to find out the density reachable set of p about ε and $MinPts$; if p is a core object, all the reachable objects of the set p form clusters; otherwise, p is not a core object, it is an outlier and should be eliminated.

The medical monitoring sensor data was expanded in the 3D numerical axes and the set of points obtained by the expansion was set to be G , and the distances of the points in the space were calculated using the Euclidean distances in Eq. (1) [12]:

$$d(P_i,Z_l)=\sqrt{\sum_{i=1}^{n}(P_{ij}-Z_{lj})^2} \tag{1}$$

where, $d(P_i,Z_l)$ represents the Euclidean distance between the coordinate value P_i of the i-th feature and the coordinate value Z_{lj} of the l-th cluster point, which is the basis for measuring the similarity between data points, P_{ij} denotes the

j-th dimensional coordinate value of the i-th feature point, and Z_{lj} denotes the j-th dimensional coordinate value of the l-th clustering point.

To cluster these points, the values of the parameters ε and $MinPts$ need to be developed first. The parameter values are influenced by the number of data points collected by the medical monitoring multi-sensory during the time and the positive proportionality that exists between them. The number of multi-sensor measurements, and hence the parameter values, are determined based on the needs of medical monitoring.

The execution flow of the DBSCAN algorithm for the medical monitoring sensor data acquisition process is as follows:

1) Arbitrarily select any point object P belonging to any data cluster in the point set G and create it as a new data cluster.

2) For all points P, if the points' number of P's neighborhood is no less than a minimum, P is served as the core point of the newly created data cluster and is compared with other data points to add the density reachable point to that cluster.

The determination of the density reachable is:

A point W is density reachable from a core point P, if there is a path from P to W, and each point on the path is a core point, and the distance between each adjacent two points on the path is within the neighborhood.

3) Loop step (2) until no points can be added to the cluster, then perform step (1) again.

4) Ends when all points in the point set belong to a particular data cluster.

The DBSCAN algorithm is shown in Fig. 1.

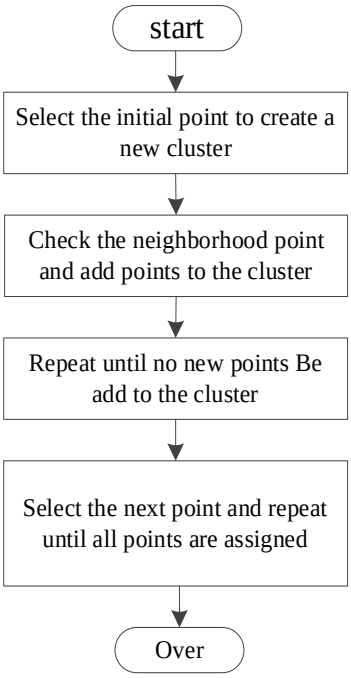

Fig. 1. The DBSCAN algorithm flowchart.

After the DBSCAN algorithm is processed, the medical surveillance multi-sensor data will exist in clusters and the number of points in the clusters a and $MinPts$ need to be compared to determine the validity of the clusters:

1) When $a < MinPts$ is present, the cluster will be considered as a cluster consisting of anomalies, which are anomalous medical monitoring sensor data, and will be excluded.

2) When $a \geq MinPts$, the cluster is composed of valid data, which will be retained to obtain the clustered collected medical monitoring multi-sensor data Q.

III. Sensor Data Denoising for Medical Monitoring

Medical monitoring data are subject to various interference during acquisition and transmission, such as electromagnetic interference, signal attenuation, and transmission failures. These factors can lead to noise in medical monitoring data, which can be mixed into the signal, masking or distorting useful information and reducing the quality and interpretability of the data, and the nonlinear relationships that exist in the noisy data can also make feature extraction difficult. Therefore, by denoising the acquired medical monitoring data, noise interference can be reduced, the accuracy and reliability of the data can be improved, it helps to highlight important features in the signal, and features related to the patient's condition can be extracted and analyzed more accurately, thus enhancing the discriminative power of the classification method.

In this study, after completing the medical monitoring data clustering, the Discrete Wavelet Transform (DWT) is used for denoising. The main advantages are: firstly, the discrete wavelet transform can effectively remove the noise in the signal and improve the quality of the data; secondly, by retaining the important features of the signal, the method can accurately denoise the signal while maintaining the detailed information of the signal as much as possible; in addition, the discrete wavelet transform can also analyse the signal at multiple scales and capture the features at different scales, which can help to understand the structure of the signal in a more comprehensive way and denoise it efficiently; Finally, the discrete wavelet transform has efficient real-time processing capability, which is suitable for scenarios that require rapid processing of medical monitoring data. The specific denoising process is shown in Fig. 2.

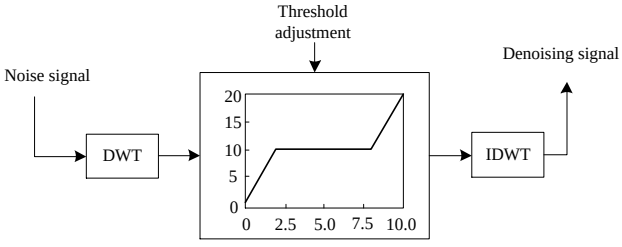

Fig. 2. Discrete wavelets transform denoising flowchart.

The algorithm flow is as follows:

Step 1. Decompose data $U(t)$ at t time into a set of wavelets with Eq. (2) [13]:

$$U(t)=\frac{1}{\sqrt{2^j}}\sum_k\sum_j DW_k^j\cdot\varphi(2^{-j}\cdot t-k) \tag{2}$$

where, φ and j denote the translation and expansion factors, respectively, and DW_k^j denotes the DWT coefficients in the wavelet basis.

Step 2. Use Multi-Resolution Analysis (MRA) data to extend analyses that provide different time and frequency resolutions at each level.

Multi-resolution analysis (MRA) provides different time and frequency resolutions at each analysis level by using the scale function $\phi(t)$ and the wavelet function $\varphi(t)$. $\phi(t)$ is the foundation of MRA, which describes the overall signal features on a coarser scale, similar to the low-frequency in the signal. It constructs the low-frequency approximation of the signal through its expansion and displacement. $\varphi(t)$ describes the local signal features on a finer scale, similar to the high frequency in the signal. It captures the signal details through expansion and displacement, that is, the high frequency in the signal at different scales.

The multi-sensor data $U(t)$ is calculated by Eq. (3), which decomposes the signal into a linear combination of the scale function and the wavelet function at different scales [14].

$$U(t)=\sum_k A_k^{j_0}\varphi_k^{j_0}(t)+\sum_{j=j_0}\sum_k D_k^j\phi_k^j(t) \tag{3}$$

where, $A_k^{j_0}$ and D_k^j denote the approximation coefficient and detail coefficient, respectively. According to the scale and wavelet bases, the equations for both can be expressed as Eq. (4) and Eq. (5) [15]:

$$A_k^{j_0}(t)=\sum_k f_k\cdot\varphi_k^{j_0}(t)=\sum_n h(n-2k)\cdot A_k^{j_0+1}(t) \tag{4}$$

$$D_k^j(t)=\sum_k f_k\cdot\phi_k^{j_0}(t)=\sum_n g(n-2k)\cdot A_k^{j+1}(t) \tag{5}$$

The wavelet decomposition tree constructed by MAR is shown in Fig. 3.

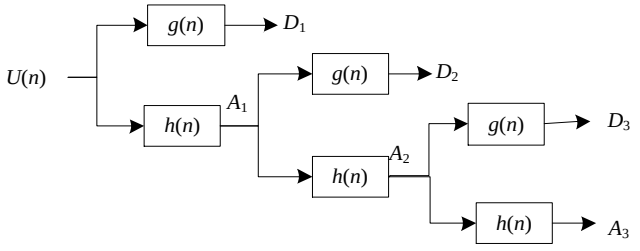

Fig. 3. Schematic diagram of MRA scale decomposition.

Step.3 Soft Threshold Calculation

In the process of quantizing the threshold λ, it is usually considered that anything less than λ is caused by noise, and anything greater than λ is caused by signals, and the soft threshold algorithm for searching towards zero according to a fixed amount is as follows in Eq. (6) [16]:

$$\hat{w}_{j,k}=\begin{cases}\mathrm{sgn}(w_{j,k})(w_{j,k}-\lambda), |w_{j,k}|\geq\lambda\\ 0, |w_{j,k}|\leq\lambda\end{cases} \tag{6}$$

In Eq. (6), λ and $\mathrm{sgn}(x)$ are defined as Eq. (7) and Eq. (8), respectively:

$$\lambda=\sigma\sqrt{2\log(n)} \tag{7}$$

$$\mathrm{sgn}(x)=\begin{cases}1, x>0\\ 0, x=0\\ -1, x<0\end{cases} \tag{8}$$

In the above equation, n denotes the sampling length of the noise signal, σ denotes the noise variance, and $w_{j,k}$ denotes the wavelet coefficients at different scales. The noise variance σ is calculated in Eq. (9) [17]:

$$\sigma=\left[\frac{median|\Lambda_{i,j}|}{0.6475}\right] \tag{9}$$

In Eq. (9), $\Lambda_{i,j}$ denotes the first layer of high-frequency coefficients of the wavelet transform of the noise-containing medical monitoring multisensory data, and $median|\Lambda_{i,j}|$ denotes the median.

Step.4 Positive and Negative Thresholding

The DWT exhibits low-frequency and high-frequency coefficients, respectively, and the data are characterized by two types of high-frequency oscillations: high-frequency vibrations with small amplitudes and low-frequency oscillations with large amplitudes. Both contain peaks and noise signals, respectively [18]. Therefore, the positive and negative thresholds are found through Eq. (10):

$$\lambda_i=\begin{cases}\lambda_+=\alpha*\max(d_i)\\ \lambda_-=\alpha*\min(d_i)\end{cases} \tag{10}$$

In Eq. (10), d_i denotes the high frequency detail coefficient of the n layer, and α denotes the empirical parameter.

The post-threshold low-frequency and high-frequency signals were reconstructed to obtain denoised medical monitoring multi-sensor data $X(t)$.

IV. Multi-Sensor Data Feature Extraction for Medical Monitoring

Due to the characteristics of high dimensionality and large-scale data, its direct use for classification processing will increase the complexity of the computational algorithm and reduce the accuracy of the algorithm. Feature extraction for the fused data after the denoising process can map the original data to a lower-dimensional feature space, reduce the data dimension, simplify the problem, improve the classification efficiency, help to extract the key information and features in the data, and strengthen the effective patterns and structures in the data. Therefore, in this study, CNN is used to extract the main features of the data to reduce the interference of unnecessary information and improve the classification ability of the data by finding meaningful features related to the classification task.

The convolutional layer is the core of the whole CNN, and it can effectively extract the sample features. The convolutional kernel traverses the whole sample by sliding to carry out feature extraction, and after the local features are extracted, the positional relationship between that part of the features and other features can be marked. CNN is non-transparent for feature extraction, and different from traditional network learning methods, the convolutional kernel operation has the feature of local perception, when a part of the features are learnt, the features will be inputted into the next mapping layer, and the mapping layer after that will continue to carry out classification learning. The samples after the convolutional layer use the spatial distribution characteristics of the surrounding samples, and the feature extraction accuracy can be obtained.

For the input medical monitoring multisensory data $X(t)$, a matrix W with a convolution kernel size $a \times b$ and bias B, the result after convolution is in Eq. (11) [19]:

$$h = g(X * W + B) \tag{11}$$

In Eq. (11), $*$ denotes the convolution operation and $g(.)$ denotes the activation function. The activation function used in this study is the ReLU function shown as Eq. (12):

$$g(x) = \max(0, x) \tag{12}$$

The sparse activation of the ReLU function can make certain neurons inactive, reduce the number of redundant connections and parameters, and improve the computational efficiency and generalization ability. The ReLU function solves the problem of gradient vanishing in traditional activation functions and avoids the performance degradation caused by the gradient being too small. In addition, the computation of the ReLU function is simple and efficient, without the need for complex exponential operations, which accelerates the training and inference process of the model. In addition, the ReLU function integrates linear and nonlinear properties, which is convenient to capture the nonlinear relationship of the data and enhances the expressive ability of the neural network. Finally, the ReLU function has no upper limit and can pass positive values to the next layer, further expanding the expressive ability of the network.

The pooling layer can effectively reduce the data dimensions, and the input of the pooling layer is the output of the convolutional layer. The output of the convolutional layer N feature maps is passed through the pooling layer to reduce the data size. In this research, the maximum pooling layer is introduced to sparsity the hidden layer data. The operation of maximization is to retain the maximum value of the pooled region, that is, to remove the non-extremely large values within the matrix, while extracting the extremely large value of the region as the representative value after pooling. The output of the pooling layer results in a substantial reduction of the network parameters, and the kernel size is generally 2×2 when maximum pooling is used, and the 2×2 size kernel enables the number of parameters to be effectively regulated, which helps to prevent the occurrence of overfitting situations and enhances the robustness of the convolutional neural network. Average pooling performs an averaging operation on the matrices in the region and uses the average value to sparse the non-overlapping target region.

The role of the Dropout layer is to drop some neurons at a certain rate during the network training process, which is a simple and effective regularization technique. However, choosing a suitable Dropout rate is critical to optimizing network performance. Higher Dropout rate causes the model to underfit because too many neurons are dropped, and difficult to learn enough information. However, a lower Dropout rate cannot effectively prevent overfitting. In addition, due to the randomness of Dropout during training, it helps the model learn more robust feature representations. Also, the difference between training and testing exists because all neurons are activated.

Therefore, it is often necessary to scale the output of all neurons to match the average level of activation at training, thus improving the generalization ability of the model. Choosing an appropriate Dropout rate is crucial to prevent model overfitting and improve generalization performance. In general, for larger networks or datasets, a lower Dropout rate (e.g. 0.2-0.5) can be chosen, as the network already has enough capacity to capture complex features, and too high a Dropout rate may limit its ability to learn. Conversely, for smaller networks or datasets, a higher Dropout rate (such as 0.5-0.8) can be selected to reduce the risk of overfitting.

The structure of the Dropout network is shown in Fig. 4.

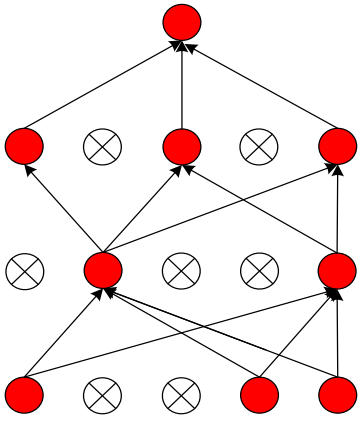

Fig. 4. Dropout network structure.

In the original network structure, all neurons need parameter training, while in the Dropout network structure, some neurons are not involved in the training, Dropout is a very effective method to suppress overfitting, and its use often brings a great improvement to the network generalization performance. The computational equation of the network using Dropout as in Eq. (13) [20]:

$$\left.\begin{aligned} r_i^l &\sim Bernoulli(p) \\ \tilde{y}^{(l)} &= r^{(l)} * y^{(l)} \\ z_i^{l+1} &= W_i^{(l+1)} \tilde{y}^{(l)} + B_i^{(l+1)} \\ y_i^{(l+1)} &= f(z_i^{(l+1)}) \end{aligned}\right\} \tag{13}$$

In Eq. (13), *Bernoulli* function to generate the probability r vector, that is, randomly generate a $[0,1]$ vector, $y_i^{(l+1)}$ is the output of the convolutional neural network for medical monitoring multi-sensor data feature extraction results.

V. Reinforcement Learning for Improved SVM Classification Models

Through feature extraction, we have extracted the most representative and distinguishing features from the original multi-sensor data. The classification algorithm can better distinguish different data types by enriching the feature space and improving classification accuracy. Many physiological signals in medical monitoring data are time-varying, and the trend of these time-varying features is very important for predicting the direction of disease development. Considering these characteristics of sensor monitoring signals, a method combining Q-learning with SVM is proposed to improve the effectiveness of SVM in multi-sensor data classification of medical monitoring.

Reinforcement learning can reduce the dimensionality of the feature space of sensor signals, improve the efficiency of the algorithm, can better capture the dynamic change law of time-varying features, and improve the classification method's ability to process time-ordered data. Therefore, in this study, the Q learning algorithm in reinforcement learning is used to improve SVM, to provide better classification effect of medical monitoring data by SVM.

SVM is a powerful supervised learning algorithm that is widely used in classification and regression tasks. In classification, the goal of SVM is to find a hyperplane that maximizes the spacing between different classes; in regression, the goal of SVM is to find a hyperplane that minimizes the distance of all data points to the hyperplane, while allowing for a certain error boundary.

In SVM regression, the objective function aims to minimize the sum of distances from all data points to the hyperplane while taking into account the error boundary. For a given training data set $\{(x_i, y_i)\}_{i=1}^n$, where x_i is the input vector and y_i is the corresponding output value, the goal of SVM regression is to minimize Eq. (14) objective function [21]:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} L_\varepsilon(y_i - (w \cdot \phi(x_i) + b)) \tag{14}$$

where, w and b are the normal vector and intercept of the hyperplane, respectively; $\phi(x_i)$ represents the mapping from input space to feature space; C represents a regularization parameter, which is used to control the balance between the complexity of the model and the training error; L is ε - insensitive loss.

In SVM, support vectors are those training samples on or within the error boundary, which have a decisive effect on the hyperplane position for their determination of the constraints on interval maximizing. The support vectors are equally important in regression because of both the definition of hyperplane location and affection of the generalization ability. By keeping the key sample points, SVM achieves better generalization performance while ensuring training errors.

Q-learning method is an important machine learning method in reinforcement learning. It is similar to other methods in reinforcement learning in that it learns the optimal policy for its dynamic system by perceiving the state information of the environment and interacting with the environment continuously, and it improves the behavior of the system by interacting with the environment through trial and error methods, and it is a kind of on-line learning method that can be applied to real-time environments. Therefore, it has been widely researched in the fields of intelligent control, machine learning, and so on.

The Q learning system views learning as a process of trial and error, and its basic model is shown in Fig. 5.

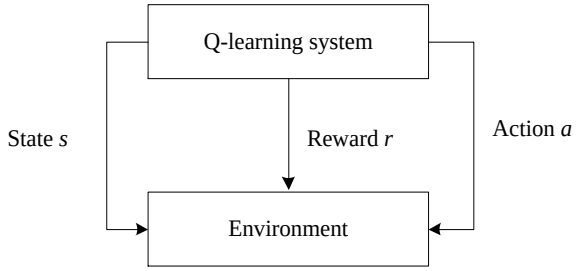

Fig. 5. Basic model of Q-learning.

In the learning process, the Q-learning system selects an action a to act on the environment, the environment receives the action and changes, and at the same time produces a reinforcement signal r which is fed back to the learning system, which then selects the next action according to the reinforcement signal and the current state of the environment, the principle of selection is to make the probability of receiving a positive reward increase. The chosen action affects not only the immediate reinforcement value, but also the next state of the environment and the final reinforcement value.

From the Q learning system, it is known that the set of states in the Q learning system is discrete and finite in general, but often in some practical situations, the set of state variables of the environment is large-scale or continuous, resulting in a difficult learning process and a slow response speed of the system. To solve this problem, this study introduces Q-learning to improve

SVM, where Q-learning is used to preprocess and extract features, and then SVM is used for classification. It not only captures the dynamic variation of time-varying features, but also reduces the complexity of SVM when processing high-dimensional data.

Q-learning improves the behavior of the system by constantly interacting with the environment through trial-and-error methods to find a strategy that maps the state of the environment into the corresponding action. This strategy of reinforcement learning inevitably increases the difficulty of learning the system, resulting in very much blind learning, and leading to longer learning times.

To overcome these problems, the features optimized by Q-learning are passed to the SVM as input to make efficient classification decisions, thus improving the overall system performance. For large-scale systems with reinforcement learning , the introduction of SVM not only reduces the complexity due to solving quadratic programming problems in standard support vector machines, but also for improving the convergence speed of large-scale reinforcement learning systems, Q-learning improved SVM as shown in Fig. 6.

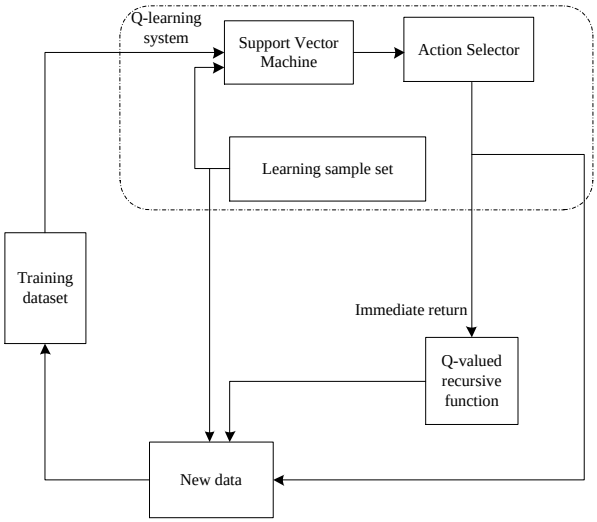

Fig. 6. Q-learning improved SVM structure.

The aim of the Q-learning algorithm shown as Eq. (15) and Eq. (16) is to find a policy π such that the value $Q^{\pi}(s)$ of each state s reaches the end at the same time, that is, it tries to find a policy $\pi : s \to a$ that is able to maximize the value of each state [22].

$$Q^{\pi}(s,a) = E\{r_1 + \gamma r_2 + \cdots + \gamma^{i+1} r_i + \cdots \mid s_0 = s, a_0 = a\} \tag{15}$$

$$Q^{\pi}(s,a) = \max_{\pi}(Q^{\pi}(s,a)) \tag{16}$$

where, r_i denotes the immediate reward at the moment t, $\gamma(\gamma \in [0,1])$ denotes the decay coefficient, and $Q^*(s,a)$ denotes the optimal value function. The corresponding optimal strategy as in Eq. (17):

$$\pi^* = \arg\max Q^*(s,a) \tag{17}$$

One of the simplest forms of Q-learning is single-step Q-learning with a modified consensus of Q-values as shown in Eq. (18) [23]:

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \beta\left[r_{t+1} + \gamma \max_a Q(s_{t+1},a) - Q(s_t,a_t)\right] \tag{18}$$

where, β denotes the learning rate and γ denotes the discount rate.

Given a training set $S = (y_i, P_i)$, where y_i denotes the input vector and P_i is the output classification objective value. Determine an optimal function $F(x)$ as the objective of the regression problem such that $F(x)$ can correctly regress the unknown input vector with the highest possible probability. In SVM, the regression function $F(x)$ has the form as in Eq. (19) [24]:

$$F(x) = \upsilon^T \delta(x) + b \tag{19}$$

where, $\delta(.)$ denotes the mapping from the input space to the feature space, υ and b denote the coefficient vector and the deviation term respectively, which are the quantities to be sought.

The unknown quantity can be determined by Eq. (20) optimization problem [25]:

$$\begin{aligned} &\min_{\upsilon,b,\varepsilon} Q(\upsilon,b,e) = \frac{1}{2}\|\upsilon\|^2 + \frac{\gamma}{2}\sum_{i=l}^{l} e_i^2 \\ &s.t. y_i = \upsilon^T \delta(x_i) + b + e_i \end{aligned} \tag{20}$$

Its Lagrangian function as in Eq. (21):

$$L(\upsilon,b,e,\beta) = Q(\upsilon,b,e) - \sum_{i=1}^{l} \beta_t\left[\upsilon^T \delta(x_i) + b + e_i - y_i\right] \tag{21}$$

The KKT condition for the above equation is shown in Eq. (22) [26]:

$$\begin{cases} \dfrac{\partial L}{\partial \upsilon} = 0 \Rightarrow \upsilon - \sum_{i=1}^{l} \beta_i \delta(x_i) = 0 \\ \dfrac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=0}^{l} \beta_i = 0 \\ \dfrac{\partial L}{\partial e_i} = 0 \Rightarrow Ce_i - \beta_i = 0 \\ \dfrac{\partial L}{\partial \beta_i} = 0 \Rightarrow \upsilon^T \delta(x_i) + b + e_i - y_i = 0 \end{cases} \tag{22}$$

Writing Eq. (22) in matrix form and eliminating υ and e, Eq. (23) is obtained:

$$\begin{bmatrix} 0 & \vec{1}^T \\ \vec{1} & \Omega+\gamma^{-1}I \end{bmatrix}\begin{bmatrix} b \\ \beta \end{bmatrix}=\begin{bmatrix} 0 \\ y \end{bmatrix} \tag{23}$$

where, $\Omega_{ij}=k(x_i,x_j)$. A regression function completes the classification of multi-sensor data for medical monitoring by solving Eq. (24):

$$F(x)=\sum_{i=1}^{l}\beta_i K(x,x_i)+b \tag{24}$$

VI. Experimental Validation

Test experiments are conducted to validate the effectiveness of the proposed reinforcement learning based improved SVM for multi-sensor data classification processing in medical monitoring.

A. Experimental Data

A schematic diagram of the medical monitoring multi-sensor arrangement is shown in Fig. 7.

Fig. 7. Schematic diagram of multi-sensor arrangement for medical monitoring.

Here, suitable types of sensors, including heart rate sensors or blood pressure sensors, are selected based on monitoring needs and placed at specific locations on the patient's body to accurately collect physiological parameters and health data. Subsequently, the data collected by the sensors is transmitted to the monitoring device or data system for processing and storage.

The multi-sensor data included ECG data (heart rate, heart rate), blood pressure data (including systolic and diastolic blood pressure), oxygen data (e.g., oxygen saturation), temperature data, respiratory data (respiratory rate), exercise data (number of steps, intensity of activity), including blood glucose data, and pulse waveform data. Examples of selected samples are shown in Table I.

TABLE I. Examples of Selected Samples

Data Type	Numerical Value
heart rate	75 bpm
pulse rate	normalcy
systolic blood pressure	120 mmHg
diastolic blood pressure	80 mmHg
oxygen saturation	98 per cent
(body) temperature	36.5°C
respiratory rate	16 times/minute
number of steps	8000 steps
Activity intensity	medium
blood sugar	5.0 mmol/L

In this setting, the parameter configurations of Q- learning and SVM model are planned in detail, and the specific parameter settings are as follows:

1) The state space of Q-learning algorithm selects parameters in multi-sensor data as characteristics, such as heart rate and systolic blood pressure. The action space includes the selection of different SVM improvement methods.

2) Setting the learning rate of Q-learning is 0.1 to control the updating speed of the Q value; the discount rate is 0.9 to balance the importance of current and future rewards; the exploration rate gradually decreases from 0.9 to 0.1 to promote the algorithm's balance between exploration and utilization.

3) Setting the initial regularization parameter C of SVM is 1.0, and the kernel function is RBF (radial basis function). The space is reserved for adjusting other parameters through cross-validation.

The above parameter configuration aims to improve the repeatability of the experiment and the reliability of the results.

B. Experimental Program

Based on the above collected data, using the signal-to-noise ratio, data feature extraction effect, and Kappa coefficient as indicators, the method of this study is compared and validated with the data classification method based on the concept of drift, and the classification method based on mathematical modelling and improved online learning.

1) Signal-to-noise ratio. The signal-to-noise ratio is the ratio of the strength of the useful signal to the strength of the noise signal and is used to measure the relative strength of the signal in relation to the noise. A high signal-to-noise ratio indicates that the useful signal is stronger relative to the noise signal, indicating a higher quality signal and clearer information.

2) Data feature extraction effect. Data feature extraction effect refers to the meaningful features that can characterize the relevant information of the signal extracted from the original data by feature extraction methods.

3) Kappa coefficient. The Kappa coefficient measures the accuracy of the classifier while considering the random factor in classification. The closer the Kappa coefficient is to 1, the better the performance of the classifier.

C. Analysis of Experimental Results

1) Signal-to-noise ratio. The signal-to-noise ratio is a measure of the relative strength between signal and noise. By performing a comparative signal-to-noise ratio test, the quality of different data can be assessed, and it can be determined which data has a higher signal quality and is more reliable. Signal-to-noise ratio is one of the key factors in signal classification and processing. Lower signal-to-noise ratios may lead to interference and distortion of the signal by noise,

reducing the performance and accuracy of the classification algorithm. By performing a signal-to-noise ratio comparison test, data with lower signal-to-noise ratios can be identified, so that methods such as appropriate preprocessing or adjusting parameters can be taken to improve classification performance. The signal-to-noise ratio results of the three methods are shown in Fig. 8.

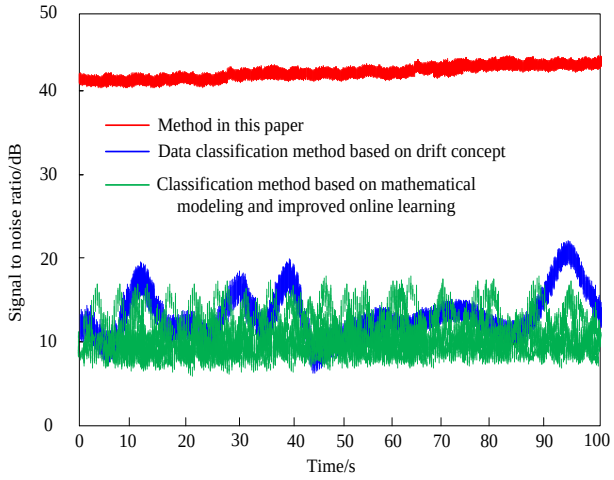

Fig. 8. Signal-to-noise ratio results.

Observing the results of the signal-to-noise ratio shown in Fig. 8, the signal-to-noise ratio of this study's method is always higher and smoother compared to the two literature comparison methods, and the signal-to-noise ratio of this study's method is always above 40 dB. On the other hand, the signal-to-noise ratio of the data classification method based on the concept of drift and the classification method based on mathematical modelling and improved online learning not only fluctuates more, but also has a lower value, with a maximum of only about 25 dB. Therefore, it shows that the method in this study can effectively remove the noise from medical monitoring multi-sensor data and improve the quality of the data.

The main reason is that the proposed method removes the noise in medical monitoring data effectively and improves the signal-to-noise ratio by combining a DBSCAN algorithm with DWT.

2) Effectiveness of data feature extraction. Feature extraction is a key step in converting raw signals into meaningful feature vectors. The performance and effectiveness of different feature extraction methods can be evaluated by verifying the feature extraction effect. Comparing the usefulness and differences of features extracted by different methods for signal classification helps to select the most suitable feature extraction method to maximize the accuracy and robustness of signal classification. The data feature extraction results of the three methods are shown in Fig. 9.

As shown in Fig. 9, the data feature extraction results of this study's method are closest to the trend of the original data, indicating that the data features extracted by this study's method are most capable of describing the original data. And the data feature extraction results of the data classification method based on the drift concept differ greatly from the original data. Therefore, it shows that the method in this study can accurately extract the features of medical monitoring multi-sensor data.

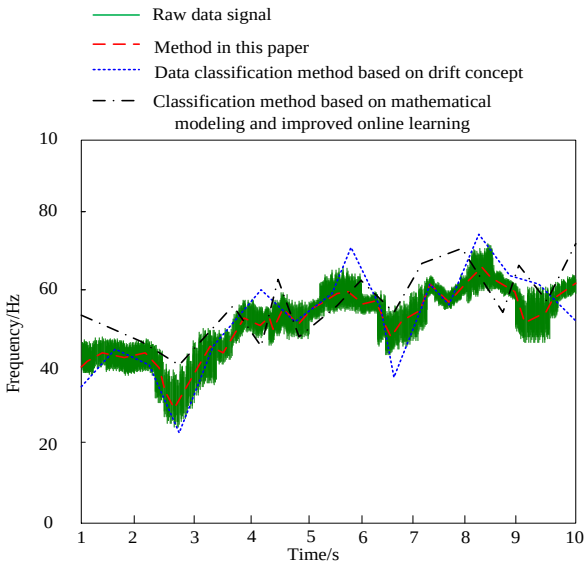

Fig. 9. Effect of data feature extraction.

This is mainly because the convolutional neural network is used in this method, which makes it more accurate in feature extraction and can capture more useful information.

3) Kappa coefficient. The accuracy and consistency of the classifiers in multi-sensor data classification can be objectively assessed by validating the classification results with Kappa coefficients. The Kappa coefficients take into account the random assignment of the classified objects, and can accurately measure the superiority of the classifiers with respect to the classification results that are only caused by the random selection. The Kappa coefficient test results of the three methods are shown in Table II.

TABLE II. KAPPA COEFFICIENT

Test serial number	Kappa coefficient		
	The method in this study	Data classification method based on the drift concept	Classification method based on mathematical modeling and improved online learning
10	0.96	0.71	0.76
20	0.98	0.68	0.74
30	0.95	0.69	0.62
40	0.97	0.72	0.64
50	0.96	0.73	0.74
60	0.96	0.67	0.65
70	0.97	0.71	0.61
80	0.98	0.70	0.63
90	0.96	0.69	0.65
100	0.95	0.66	0.71

From the data in the table, it can be observed that the Kappa coefficient of this study's method stays at a high level with small changes in many tests. On the other hand, the Kappa coefficients of the other two methods have large fluctuations and their performance is not stable enough. For example, the Kappa coefficient of this study's method stays above 0.95 under different test numbers, while the Kappa coefficients of the other two methods fluctuate in a wide range, from 0.62 to 0.76. The Kappa coefficient of this study's method is significantly higher than that of the other two methods under most of the test sequence numbers, especially the highest Kappa coefficient is achieved at many nodes such as test sequence numbers 20, 40 and 80. This indicates that the method in this study has a significant advantage in classification accuracy and can classify the data more reliably for judgement.

This is because the improvement of Q-learning on SVM makes the model more adaptable to complex and changeable medical data through automatic exploration and optimization of parameters, thus improving the accuracy and robustness of the classification.

VII. Conclusion

In this study, a method combining Q-learning and SVM was proposed to deal with the problem of multi-sensor data classification in medical monitoring. Through experimental validation, the new method exhibited better performance relative to the traditional approach. Specifically, DBSCAN clustering and DWT were used to preprocess data, which effectively retained key features and reduced noise interference. Then, CNN was used to deeply mine signal features. The Q-learning algorithm was used to optimize SVM model and realize efficient mapping of features to classification output.

Experimental results show that the proposed method not only reaches a high level of 98% on classification accuracy, but also observes significant progress in terms of classification accuracy, signal-to-noise ratio, and feature extraction effectiveness. In addition, this study explores the role of reinforcement learning in the process of feature selection and model optimization, and it is found that the reinforcement learning algorithm can be used to better guide the model for tuning and improve the effectiveness of the classification task.

However, this study still has certain limitations, and the robustness of feature extraction in current methods needs to be further improved when dealing with sudden noise or multi-source heterogeneous sensor data. And the convergence speed of the Q-learning algorithm in high-dimensional state space is slow, which may affect the application efficiency of real-time monitoring scenarios.

In the future, this study will further develop a distributed reinforcement learning framework to optimize the parallel processing efficiency of multi-sensor data, in order to meet the real-time monitoring needs of clinical practice. For example, introducing Deep Q-Network (DQN) combined with Convolutional Neural Network to process high-dimensional sensor data, utilizing experience replay and target network mechanisms to enhance algorithm stability. And adopting the Actor Critic framework, while optimizing the strategy function and value function, to more efficiently handle the temporal dynamic characteristics of medical monitoring data.

Acknowledgment

The authors are thankful for the support from the Key Scientific Research Project of Higher Education Institutions in Henan Province (grant no.24B 510006).

References

[1] Kapoor B, Nagpal B, Alharbi M. Secured healthcare monitoring for remote patient using energy-efficient IoT sensors. Computers & Electrical Engineering, 2023, 106: 108585.

[2] Wang S, Nayak D R, Guttery D S, et al. COVID-19 classification by CCSHNet with deep fusion using transfer learning and discriminant correlation analysis. Information Fusion, 2021, 68: 131-148

[3] Kilinc D, Gel E S, Sir M Y, et al. Statistical characterization of patient response to offered access delays using healthcare transactional data. Naval research logistics, 2022, 69(7): 974-995.

[4] Xiao Y, Yin H, Duan T, et al. An Intelligent prediction model for UCG state based on dual-source LSTM. International Journal of Machine Learning and Cybernetics, 2021, 12: 3169-317

[5] Ping S , Bishop J L .Evaluation of a novel salaried medical officer position on service provision and performance at a rural health service: An exploratory mixed-methods study. Australian Journal of Rural Health, 2022, 30(1):65-74.

[6] Liu X, He J, Song L, et al. Medical Image Classification based on an Adaptive Size Deep Learning Model. ACM Transactions on Multimedia Computing Communications and Applications, 2021, 17(3S): 102.

[7] Hasib K M, Towhid N A, Islam M R. HSDLM: A Hybrid Sampling With Deep Learning Method for Imbalanced Data Classification. International journal of cloud applications and computing, 2021, 11(4): 1-13.

[8] Singh M N, Khaiyum S. Enhanced data stream classification by optimized weight updated meta-learning: continuous learning-based on concept-drift. International journal of web information systems, 2021, 17(6): 645-668.

[9] Kenger M N , Ozceylan E . A hybrid approach based on mathematical modelling and improved online learning algorithm for data classification. Expert Systems with Applications, 2023, 218: 119607.

[10] Zhai J, Qi J, Shen C. Binary imbalanced data classification based on diversity oversampling by generative models. Information Sciences, 2022, 585: 313-343.

[11] Liang X, Yu Q, Zhang K, et al. LapRamp: a noise resistant classification algorithm based on manifold regularization. Applied Intelligence, 2023, 53(20): 23797-23811.

[12] Jain P, Bajpai M S, Pamula R. A Modified DBSCAN Algorithm for Anomaly Detection in Time-series Data with Seasonality. The international Arab journal of information technology, 2022, 19(1): 23-28.

[13] Batzelis E, José M B, Toledo F J, et al. Noise-Scaled Euclidean Distance: A Metric for Maximum Likelihood Estimation of the PV Model Parameters. IEEE journal of photovoltaics, 2022, 12(3): 815-826.

[14] Liu S, Lu M, Liu G. A Novel Distance Metric: Generalized Relative Entropy, Entropy, 2017, 19(6): 269

[15] Bittner D , Engel M , Wohlmuth B ,et al. Temporal Scale-Dependent Sensitivity Analysis for Hydrological Model Parameters Using the Discrete Wavelet Transform and Active Subspaces. Water Resources Research, 2021, 57(10), 2020WR028511.

[16] Fu W, Luo Z, Liu S, et al. Spatiotemporal correlation based self-adaptive pose estimation in complex scenes, Digital Communications and Networks, 2024, online first, 10.1016/j.dcan.2024.03.007

[17] Jahanbanifard M , Price E , Gonzalez B A ,et al.A novel method to analyse DART TOFMS spectra based on Convolutional Neural Networks: A case study on methanol extracts of wool fibres from endangered camelids[J].International journal of mass spectrometry, 2023. DOI:10.1016/j.ijms.2023.117050.

[18] Zamli K Z , Din F , Alhadawi H S . Exploring a Q-learning-based chaotic naked mole rat algorithm for S-box construction and

optimization[J].Neural computing & applications, 2023, 35(14):10449-10471.

[19] Akin E , Demir K , Yetgin H . Multiagent Q-learning based UAV trajectory planning for effective situational awareness[J].Turkish Journal of Electrical Engineering and Computer Sciences, 2021, 29(5).DOI:10.3906/ELK-2012-41.

[20] Du H, Yu S. Dynamic Obstacle Avoidance for Service Robots Based on Spatio-Temporal Graph Attention Network. Computer Engineering, 2024, 50(2): 105-112.

[21] Delilbasic A , Saux B L , Riedel M ,et al.A Single-Step Multiclass SVM based on Quantum Annealing for Remote Sensing Data Classification[J].ArXiv, 2023, 17: 1434-1445.

[22] Delilbasic A, Le Saux B, Riedel M ,et al.A single-step multiclass SVM based on quantum annealing for remote sensing data classification. IEEE journal of selected topics in applied earth observations and remote sensing. 2023, 28(17): 1434-1445.

[23] Lian B, Xue W, Lewis FL,et al. Inverse Q-learning using input–output data. IEEE Transactions on Cybernetics. 2023, 54(2):728-738.

[24] Song S, Pan L, Liu S. A Q-learning based auto-scaling approach for provisioning big data analysis services in cloud environments. Future Generation Computer Systems. 2024, 154: 140-150.

[25] Wang J, Mi X, Shen H ,et al. Optimal Control for Interconnected Multi-Area Power Systems With Unknown Dynamics: An Off-Policy Q-Learning Method. IEEE Transactions on Circuits and Systems II: Express Briefs. 2023, 71(5): 2849-2853.

[26] Charvadeh YK, Yi GY. Understanding Effective Virus Control Policies for Covid-19 with the Q-learning Method. Statistics in Biosciences. 2024, 16(1): 265-289.