applied sciences

MDPI

Article

One-Class SVM Model-Based Tunnel Personnel Safety Detection Technology

Guosheng Huang *, Jinchuan Chen and Lei Liu

College of Automation, Guangdong University of Technology, Guangzhou 510006, China
* Correspondence: 2112004086@mail2.gdut.edu.cn

Abstract: The judgment of tunnel personnel's safety status mainly requires the collection of construction personnel's physical signs and cave environment data, and the early warning of abnormal status usually requires professional staff to make rapid judgments in a short time, which is costly and inefficient in terms of operation and maintenance. A single-classification support vector machine-based personnel safety status detection and early warning model is proposed to address this phenomenon. First, by deploying sensor devices at the site, we obtain data on the safety state of an actual tunnel construction scene and construct an OCSVM model for abnormal state prediction. Then the model is retained for early warning state testing, collecting relevant environmental data as well as construction personnel's physical signs data from engineering examples. Finally, we conduct horizontal different parameter model experiments and vertical different early warning state proportional data experiments to evaluate the performance of the model for personnel information security state judgment. The experimental results show that the accuracy rate of personnel security status early warning reaches more than 90%. In particular, it provides a more efficient detection means for the judgment of personnel security status.

Keywords: single-classification support vector machine; personnel safety status detection; tunnel construction; OCSVM model

Citation: Huang, G.; Chen, J.; Liu, L. One-Class SVM Model-Based Tunnel Personnel Safety Detection Technology. *Appl. Sci.* **2023**, *13*, 1734. https://doi.org/10.3390/app13031734

Academic Editor: Krzysztof Koszela

Received: 1 December 2022
Revised: 19 January 2023
Accepted: 20 January 2023
Published: 29 January 2023

1. Introduction

In recent years, the issue of early warnings of staff safety status as well as alarms during tunnel construction has become a research hotspot, and is one of the important aspects to ensure the safe development of a project. Taking into account the special characteristics of tunneling projects, some studies have established a big data platform for tunnel boring machine engineering management integrating intelligent monitoring, comprehensive analysis, collaborative management and big data application, which provides a reference for the design and development of big data platforms in the field of tunnel construction from the management and technical perspectives, and improves the efficiency and intelligent management of tunnel boring machine construction [1]. However, this big data platform still requires professional operation and maintenance personnel to rely on their work experience to carry out human–machine interaction, which is not conducive to automated information control. In order to promote the intelligent development of information management, Yanru Chen coupled a genetic algorithm and a limit-based learning machine, and used the genetic algorithm to optimize the initial weights and thresholds of the limit-based learning machine, thus greatly improving the accuracy of prediction [2]. This combination could better predict the results under different working conditions, which is the goal of the combination of intelligent algorithms in engineering safety assurance.

Along with the development of machine learning and deep learning theory research, there are a wide range of applications in various areas of engineering, and there are many algorithmic techniques in machine learning methods that can be used for safety detection, among which the single-classification support vector machine has a strong classification ability and simple model principle. Huiling Cao et al. applied a single-classification support vector machine to fault detection of aero-engines to identify fault data accurately and quickly and find abnormal engine operation status in time [3]. Wang Cheng used a two-threshold OCSVM algorithm with a two-level decision boundary for wake detection [4], and the experimental results showed that the improved algorithm improved the detection accuracy at different signal-to-noise ratios. Yufeng Gao proposed an equipment fault-detection method based on a single-classification support vector machine (OCSVM) [5], and the accuracy of the established prediction model for gas outflow from coal mining workings was higher than the prediction accuracy when applying a BP neural network. Hao-Nan Zheng's distribution network communication network traffic anomaly detection method using a single-classification support vector machine (OCSVM) [6] greatly improved the efficiency of anomalous-state detection. In the field of deep learning, in order to effectively solve the traditional construction site safety production management problem in China and reduce the problem of major casualties caused by personal negligence of construction workers, such as not wearing helmets, Qin Jia and other researchers proposed a deep-learning-based helmet-wearing detection and tracking technology [7]. The combination of other deep learning technologies in the field of engineering and construction, using deep learning algorithms to automatically train a set of target-detection models and thus improve the efficiency, is a developing trend [8].

In recent years, there have been numerous studies on safety management and risk assessment. Among these, Fang Yan et al. [9] proposed a new risk-calculation method called risk gridding (RM) for three-dimensional risk assessment based on field theory, which laid the theoretical foundation for further development of three-dimensional risk assessment based on field theory. Safety management and risk assessment are a type of system engineering, which is extensive, difficult to manage, and comprehensive [10], and it is difficult to evaluate it scientifically. Peng et al. [11] considered that laboratory safety management evaluation is essentially a multi-indicator decision problem, and Zhang YJ et al. [12] constructed an evaluation method that combines the Delphi method and hierarchical analysis. Tang et al. [13] proposed an evaluation index system for university laboratory fire safety based on "human, machine, environment and management", used queuing theory to assign weights to evaluation indexes and established a comprehensive evaluation model of physical elements. Ren Ying [14] used hierarchical analysis and a fuzzy weighted average operator to make a fuzzy comprehensive evaluation of laboratory safety management.

A series of scientific achievements have been made in research on construction site personnel localization, personnel attributes and component identification. Among them, the SVM algorithm has good feasibility and generalizability when applied to tunnel construction site safety management. In the field of tunnelling, the monitoring of personnel safeguards based on machine learning methods is widely used, and current research has focused on the wearing of personal protection equipment, simple personnel behavior, component breakage and component deformation degree recognition. These applications can solve specific safety management problems, such as helmet wearing and component crack detection. However, as a complex systemic project, tunnel construction safety management requires multiple considerations to meet various safety management needs. Based on this, the current research is somewhat one-sided and should be conducted systematically and thoroughly for the characteristics of construction safety management. For personnel safety, it is more direct and accurate to combine the characteristics of the internal environment of the tunnel to warn about the abnormal state of personnel signs. Since SVM methods have been maturely applied in research on machinery and aviation engineering, tunnel personnel safety detection belongs to a typical anomaly detection problem with a data imbalance, where anomalous data fluctuate in the case of training models with normal data. The OCSVM model can effectively solve the data imbalance problem and accurately predict anomalous states. The contributions of this paper are as follows:

- This paper establishes a set of tunnel construction personnel safety status detection and early warning models based on the OCSVM algorithm.
- This paper provides a dataset in the field of tunnel safety personnel status detection, which can be used for relevant research in this field.
- This paper reports on ablation comparison experiments conducted to corroborate the accuracy of the model on datasets with different distribution features and to consider the influence of the abnormal data proportion on the accuracy of the model. The accuracy of the model under four feature cases was also compared horizontally, and the performance of the model was evaluated comprehensively.

The identification of tunnel construction personnel's safety status requires fast and efficient detection of personnel vital signs, and timely feedback on the health status and location of abnormal personnel, so that when an accident occurs, we can efficiently, quickly and accurately obtain the internal environment of the tunnel, the status of personnel and their location, which are beneficial to project managers and rescue workers to select a rescue plan to protect personnel's lives. Therefore, this article takes the analysis of the factors affecting the signs of tunnel construction personnel and environmental factors as the entry point, and combines the OCSVM safety warning model for intelligent safety management to realize the information control of intelligent site safety.

2. Introduction of One-Class SVM Algorithm

This subsection introduces the basic principle of the OCSVM algorithm (one-class support vector machine) [15]. We train the OCSVM on normal data samples and establish an optimal hyperplane in the feature space, in accordance with the principle of maximizing the interval value to achieve the separation of the training samples and their origin. A schematic diagram of the algorithm is shown in Figure 1.

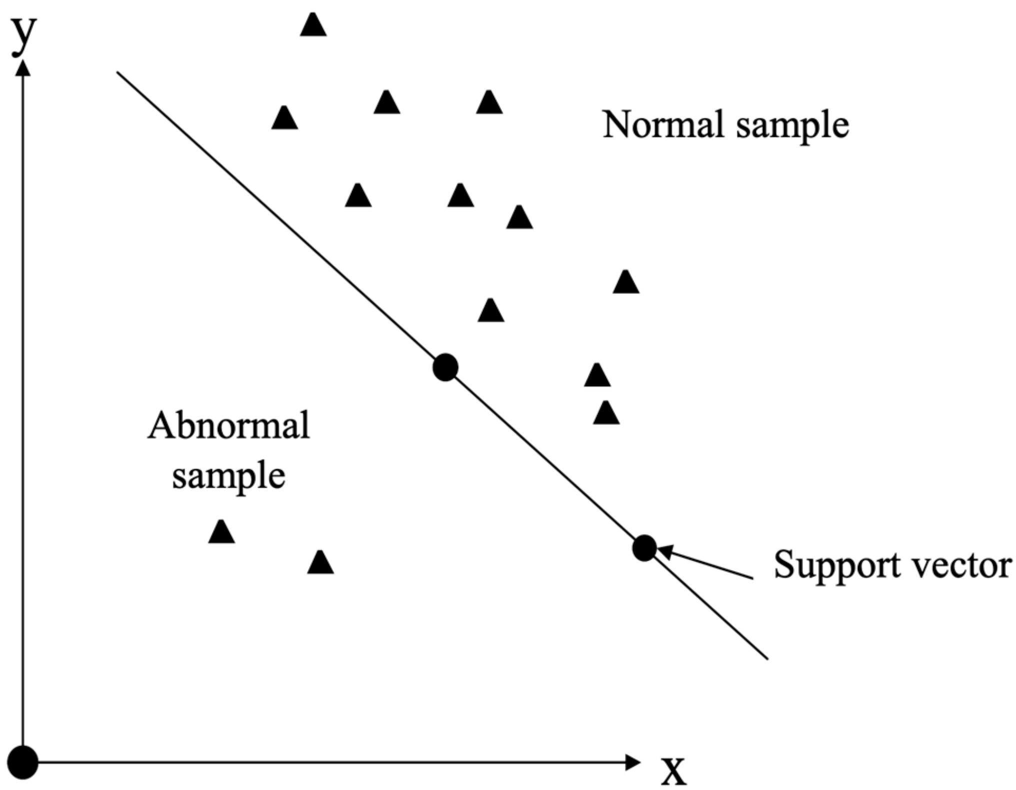


Figure 1. Schematic diagram of support vector classifier.

We let the dataset of normal samples be $X = [x_1, x_2, \ldots, x_n]^T \in R^{N\times L}$, and the target decision hyperplane be $f(x) = \omega\phi(x) - \rho = 0$. We construct and solve the optimization problem:

$$\min_{\omega\in F,\xi\in R^N,\rho\in R} \frac{1}{2} \parallel \omega \parallel^2 + \frac{1}{N_v}\sum_{i=1}^{N}\xi_i - \rho \quad (1)$$

$$s.t.\ \omega\phi(x_i) \geq \rho - \xi_i, \xi_i \geq 0 \quad (2)$$

In the above equation, N is the length of the dataset used for training; v is the regularization parameter; ξ_i is the slack variable corresponding to each dataset; ω and ρ are the decision planes that can be decided with participation; and ϕ is the way the data are spatially mapped [16]. Next, the OCSVM model is solved by introducing Lagrange multipliers $\alpha_i \geq 0$ and $\beta_i \geq 0$ and solving the Lagrange equations [17] as follows.

$$L(\omega, \xi, \rho, \alpha, \beta) = \frac{1}{2} \parallel \omega \parallel^2 + \frac{1}{N_v} \sum_{i=1}^{N} \xi_i - \rho - \sum_{i=1}^{N} \alpha_i[\omega\phi(x_i) - \rho + \xi_i] - \sum_{i=1}^{N} \beta_i\xi_i \quad (3)$$

The pairwise form of the optimization problem can be obtained by partial differentiation of the variables in the above equations [11]:

$$\min_{\alpha} \alpha^T H\alpha \quad (4)$$

$$s.t.\ \leq \alpha_i \leq \frac{1}{N_v}, \sum_{i=1}^{N} \alpha_i = 1 \quad (5)$$

H is the kernel matrix consisting of H_{ij}, and H_{ij} can be expressed as:

$$H_{ij} = K(x_i, x_j) = \phi(x_i)\phi(x_j) \quad (6)$$

$K(x_i, x_j)$ is the kernel function, which in this paper are linear and Gaussian kernel functions [17]. As for the RBF kernel function, there is only one parameter σ to be adjusted, which directly affects the width of the RBF kernel function, and the calculation formula is shown in Equation (7).

$$K(x_i, x_j) = e^{\frac{-\|x_i - x_j\|^2}{2\sigma^2}} \quad (7)$$

By solving the above quadratic programming problem, α can be solved, whereupon ω and ρ can be computed separately [18]:

$$\omega = \sum_{i=1}^{N} \alpha_i\phi(x_i) \quad (8)$$

$$\rho = \sum_{j=1}^{N} \alpha_j K(x_i, x_j) \quad (9)$$

A decisional hyperplane in the feature space can be found from the solved ω and ρ.

For the training set $Z = [z_1, \cdots, z_M]^T \epsilon R^{M\times L}$, a decision function $F(z_k)(k = 0, 1, \cdots, M)$ based on the Euclidean distance is established for classifying the test sample z_k [19].

$$F(z_k) = sign[\omega\ \phi(z_k) - \rho] \quad (10)$$

The decision function $F(z_k)$ of the test sample z_k is calculated by determining the position of z_k in relation to the decision hyperplane. If $F(z_k)$ = +1, z_k falls within the decision hyperplane and is classified as a positive class, which can be judged as a normal sample; if $F(z_k)$ = -1, z_k falls outside the decision hyperplane and is classified as a negative class, which can be judged as an abnormal sample [20].

The flow chart of the algorithm is shown in Figure 2.

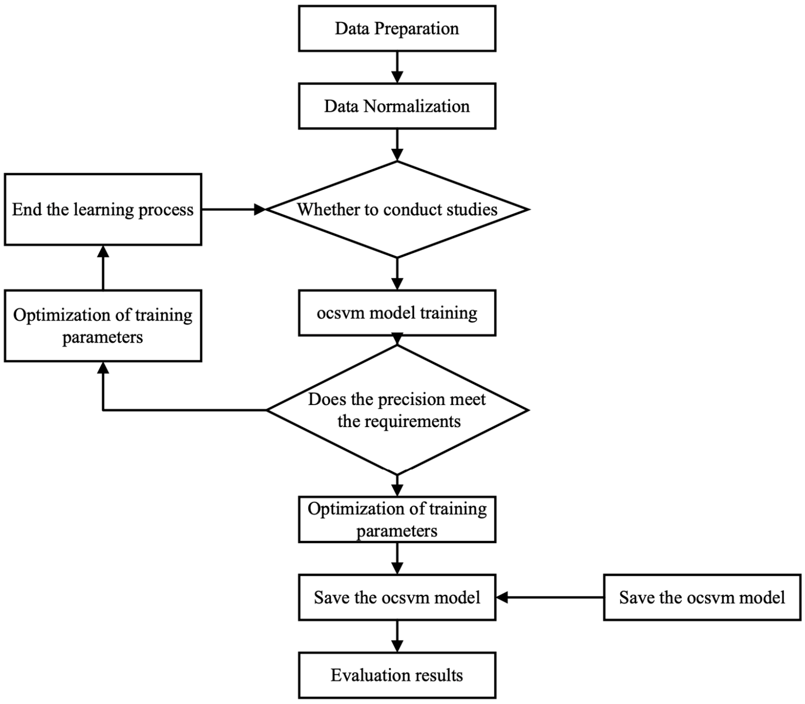


Figure 2. OCSVM model algorithm flow chart.

3. Tunnel Operator Safety Status Prediction and Early Warning Workflow

The workflow of OCSVM-based workforce safety status prediction and early warning is shown in Figure 3. It is mainly divided into three modules: data collection, model building and personnel safety detection.

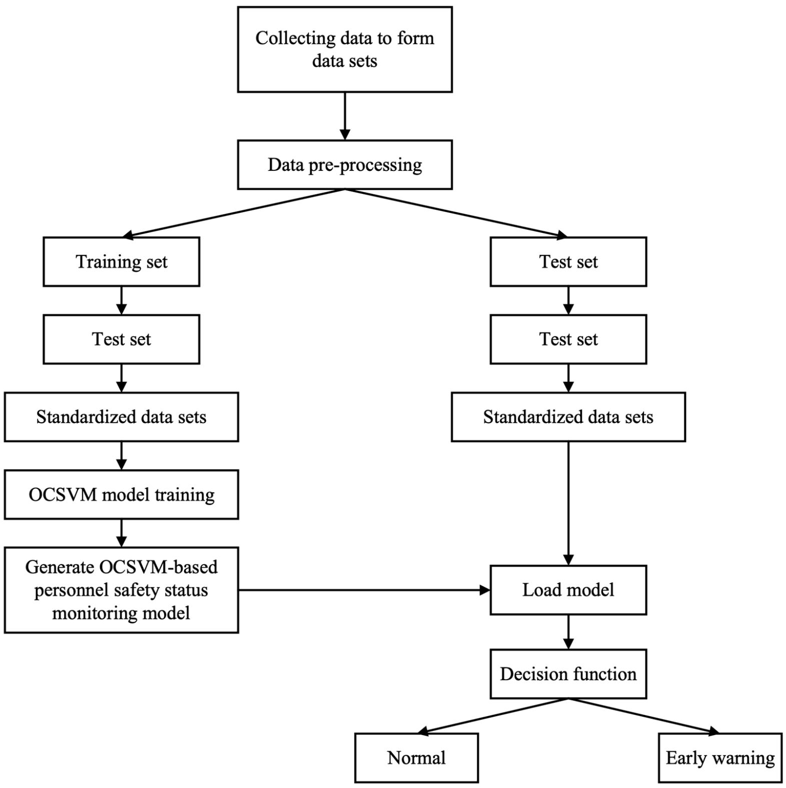


Figure 3. Tunnel work safety inspection and alarm process.

Data collection: The data sources are sent to the upper end for collation by deploying relevant sensor devices at the work site through data transmission, and the process is shown in Figure 4. It is mainly divided into the data collection layer, data transmission layer and data application layer. The main roles are described below.

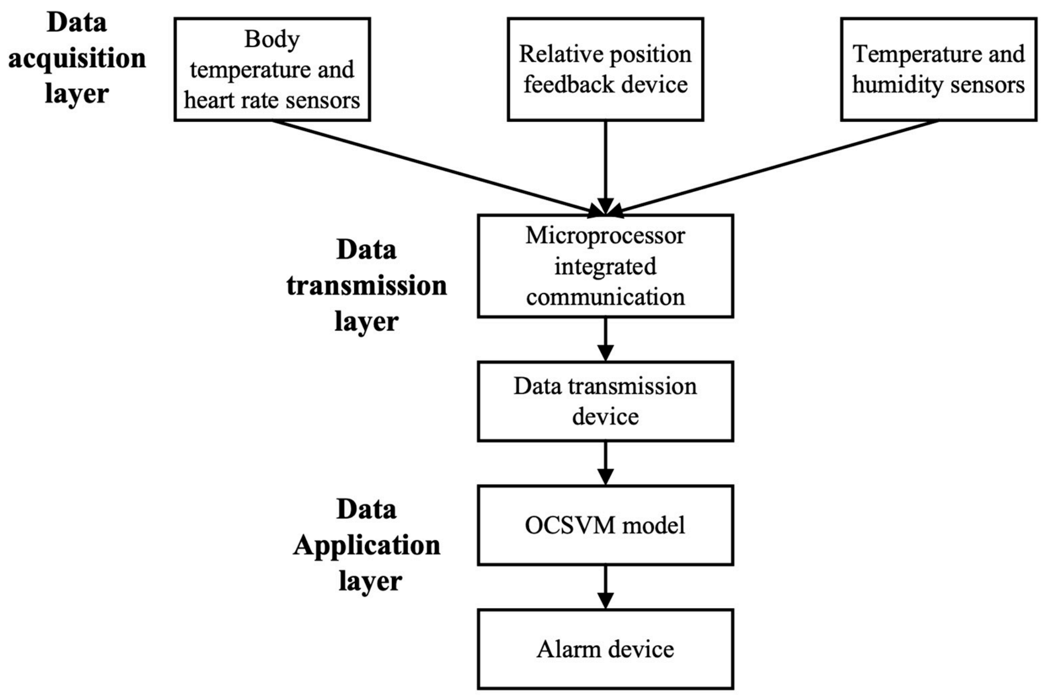


Figure 4. Data collection hardware framework.

- Data acquisition layer: The discrete data required for the experiment are obtained through the body temperature and heart rate sensors, relative position feedback devices, and temperature and humidity sensors. The data are sent to the data transmission layer to complete the transmission.
- Data transmission layer: This layer mainly consists of a microprocessor-integrated communication and data transmission device. The former has the advantage of higher integration, which can increase the rate of data transmission, and then the data are sent to the upper computer model through WiFi, Zigbee, etc.
- Data application layer: The OCSVM model acquires the data sent over and provides real-time feedback to complete the detection of staff safety status as well as early warning functions. The model can also be updated in real time using the data transmitted over.

Model building: The process is shown in Table 1.

Table 1. Model building steps.

Steps	Measures
1	Normalize the training set $X \in R$ before modeling, as described below
2	Find the model optimization parameters on the normalized training set $X \in R$
3	Build an OCSVM model for personnel safety detection based on the found optimization parameters [21]

Personnel security status detection: The basic process is shown in Table 2.

In the OCSVM model built in this paper, there are two types of model outputs, i.e., normal and abnormal cases. In terms of the output warning value, this paper sets a warning threshold of 0.78 (the threshold is set according to the statistical analysis of tunnel construction monitoring data and relevant expert experience, and can be further optimized according to the actual situation). If the output accuracy value exceeds 0.78, it indicates an abnormal situation, and the closer to 1, the more obvious the abnormal situation. That is, an output value in the range of (0–0.78) means that the situation in the tunnel is normal; (0.78–1) means that there is an abnormal situation in the tunnel.

Table 2. Personnel safety status testing.

Steps	Measures
1	Staff information data were introduced as training variables in the test set $Z_0 \in R^{M\times L}$. The test set data were processed using the same mapping rules as those normalized for the training set [22]
2	The standardized test set $Z \in R^{M\times L}$ is processed using the established OCSVM model
3	Personnel safety abnormality prediction and alarms are based on the obtained measured data decision function $F(z_k)$

4. Dataset Preparation

4.1. Data Description

As the operating environment in the tunnel is relatively closed and the site environment is complex, the temperature and humidity of the construction personnel in the cave at different locations and at different times are analyzed to screen suitable feature data for the subsequent construction of the OCSVM-based personnel safety state identification model to achieve the effect of safety warning.

(1) In the actual operation process, if there are symptoms such as fatigue, weakness and fuzzy consciousness, it will pose a threat to the safety of construction personnel. Body temperature and heart rate can directly respond to these physical signs. Body temperature of all personnel entering the construction site should be tested for strict prevention and control of epidemics. Personnel with a body temperature $\geqq$ 37.3 °C should be sent to the isolation observation quarters in a timely manner for further isolation and observation. Because the heart rate parameters are very sensitive to temperature changes, it is easy to measure and calculate, and the measurement of consumption is less time-consuming. The normal heart rate of adults is 60~100 times/minute. During operation in a closed space, if the heart rate of the staff exceeds the normal value range, the work needs to be stopped immediately to prevent the staff's body and function from becoming impaired, thus causing serious work accidents.

(2) The factors affecting temperature and humidity inside the cave are very complex, and include mechanical ventilation, natural air, groundwater, temperature and humidity outside the tunnel (which may vary daily and seasonally), heat released from construction machinery, geothermal heat, air pressure, wind direction and wind speed changes [23]. Temperature and humidity can respond not only to the operating conditions of the engineering and construction personnel, but also to the safety of the tunnel structure.

- When the humidity inside the cave is too high, the air contains more water vapor, the sweat discharged by the human body does not easily evaporate and the body does not dissipate heat smoothly, resulting in an increase in body temperature. When the temperature inside the cave decreases to a certain range, it will directly affect the physiological and mental state of the human body.
- The low-temperature situation will make the human immune system decline and cause a series of diseases, such as colds, asthma, etc. The effect of a high humidity environment makes the friction between the sweat and clothing in our body increase. High temperature and humidity have a negative impact on the operators in the tunnel, affecting the efficiency of work, possibly even endangering life and also increasing the incidence of safety accidents [24].

(3) During the tunnel construction process, the relative positions of the construction personnel have an important impact on the safety of personnel, as different construction positions may produce different degrees of danger and different search and rescue difficulties. Tunnel construction is generally divided into the cavity entrance area, second lining construction completion area, second lining construction area, waterproofing construction area, back arch/leveling layer construction area, excavation area and initial support area.

Each area is occupied by construction personnel, and the location of each person in the tunnel needs to be known. The installation of the personnel positioning device requires the determination of the coordinates of each base station, using the cave entrance as the starting point. The position of the personnel in the tunnel allows the calculation of the distance between the personnel and the palm face. Locating the distance of the personnel inside the cave from the cave entrance provides real-time statistics on the total number of personnel in the tunnel, the distribution of personnel in each area and other data. When the safety of personnel is in danger, the position of personnel from the cave entrance can be accurately determined to facilitate rescue.

In this paper, the concept of relative position is defined formally, the location of construction personnel is regionalized, the distance from the cave entrance is identified as the absolute position, the decision of the reference position is based on the construction situation on site and the size of the reference position is defined as 50 m, as shown in Figure 5. Thus, the relative position value = absolute position/50, rounded to an integer, and Table 3 describes its relevance.

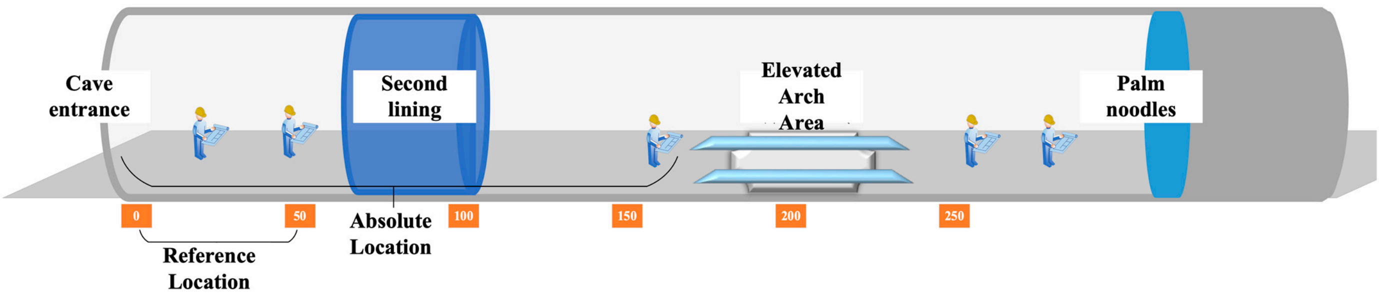


Figure 5. Relative position abstraction diagram.

Table 3. Relative position rules.

No.	Absolute Position (m)	Relative Position
1	0–50	1
2	51–100	2
3	101–150	3
4	151–200	4
5	201–250	5
……	……	……

After the formal definition of relative positions, the information search framework is shown in Figure 6.

The information collection consists of three main aspects, namely the internal environment of the tunnel, the location information and the construction personnel's physical signs. The main measurements are the temperature and humidity of the internal tunnel environment, the body temperature and heart rate of the construction personnel and the location information obtained from the relative position defined above. Thus the evaluation parameters are the human skin temperature (including the chest), heart rate, the ambient temperature inside the tunnel, the ambient humidity, and the relative position of the human body, respectively. The data in this paper are obtained from the Kuitun River

Diversion Tunnel Project in Xinjiang. The second, third and fourth standard diversion tunnel project, with a total length of 8195 km, started on 28 October 2020 and is scheduled to be completed on 1 July 2023. The basic information of the tunnel project is shown in Table 4. For the description of the subsequent experimental process, the data represented by the second, third and fourth bids in the table are defined as dataset 1, dataset 2 and dataset 3 respectively in this paper. We collected 5000 sets of data for each dataset, and each set included five variables, namely human body temperature, human heart rate, cave ambient temperature, cave ambient humidity and relative position.

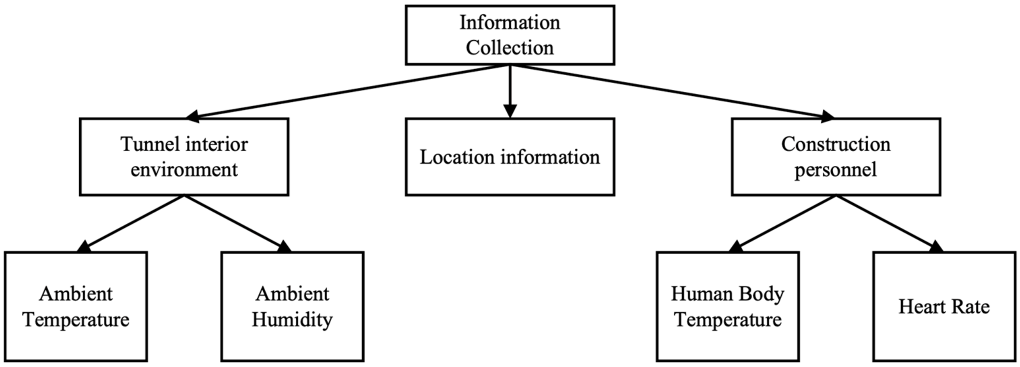


Figure 6. Schematic diagram of information collection on construction personnel.

Table 4. Basic information of tunnel engineering.

Marking Section	Name of Branch Hole	Length of Support Hole (m)	Tunnel Length (m)	Upper (Lower) Tour Chief (m)	Accumulation of Main Hole Feed (m)	Distance from the Cave Entrance (m)	Relative Position Maximum
Second bid: 2 + 095~5 + 515	1# branch hole (3 + 880)	325	3420	Upper: 1785	49	374	8
				Lower: 1635	61	386	8
Third bid: 5 + 515~8 + 205	2# branch hole (7 + 330)	494	2690	Upper: 1815	167	661	14
				Lower: 875	100	594	12
Fourth bid: 8 + 205~10 + 290	2# branch hole (9 + 400)	380	2085	Upper: 1175	137	517	11
				Lower: 890	165	545	11

4.2. Data Pre-Processing

This subsection focuses on the pre-processing of the features of the collected data. First, the descriptions of the personnel sign data and the internal environment data are shown in Table 5.

Table 5. Descriptions of personnel safety testing data.

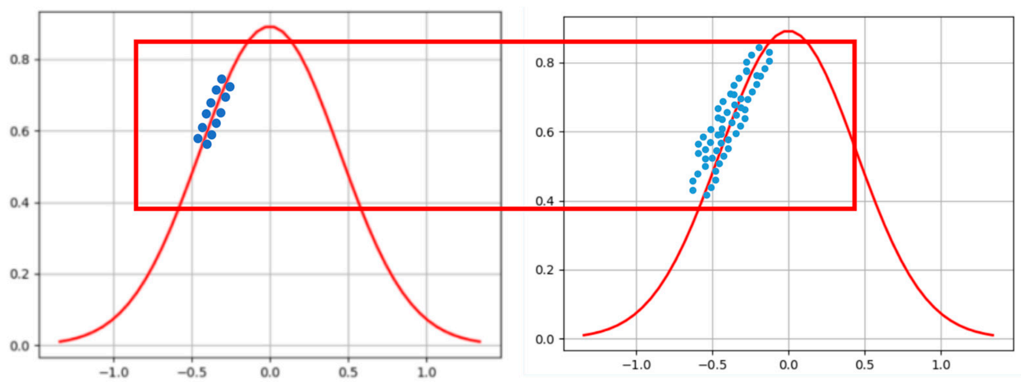
Data Name	Unit	Remarks
Ambient Temperature	Celsius (°C)	Temperature inside the cave in the area of the operator
Ambient Humidity	%	Humidity in the cave in the area where the operator is located
Relative Location	1	Relative position correspondence rule (see Figure 5, Table 3)
Body Temperature	Celsius (°C)	Skin surface body temperature sensor collection
Heart Rate	BPM	Heart rate sensor collection

In this paper, three sets of real-world engineering data were used for training, which resulted in a tunnel personnel safety detection alarm model. The statistical features of the data are shown in Table 6.

Table 6. Dataset characterization.

Data	Sensor Types	Statistical Characteristics			Reference Value	Early Warning Ratio
		Maximum Value	Minimum Value	Standard Deviation		
Dataset 1	Body Temperature Sensors	37.0	34.9	0.66	36–37.2 °C	0.1
	Heart Rate Sensor	140	50	15.08	60–100 times/min	
	Ambient Temperature Sensor	16.2	4.8	2.5859	5~20 °C	
	Ambient Humidity Sensors	83	45	11.0312	45~65%	
	Relative Position Sensors	8	1	1.7023		
Dataset 2	Body Temperature Sensors	37.2	35.1	0.71	36–37.2 °C	0.098
	Heart Rate Sensor	146	50	15.46	60–100 times/min	
	Ambient Temperature Sensor	21.5	8.2	2.3316	5~20 °C	
	Ambient Humidity Sensors	78	50	8.0320	45~65%	
	Relative Position Sensors	14	1	3.2348		
Dataset 3	Body Temperature Sensors	37.1	35	0.65	36–37.2 °C	0.08
	Heart Rate Sensor	144	50	14.25	60–100 times/min	
	Ambient Temperature Sensor	20.6	7.3	2.6042	5~20 °C	
	Ambient Humidity Sensors	76	48	11.4427	45~65%	
	Relative Position Sensors	11	1	2.5089		

On the basis of the collected data, the actual personnel security status is judged and manual alarm tagging is performed, specifically tagging for two classes, i.e., normal status and alarm status. Due to the small amount of tagging data and consequent need for a data enhancement process, the specific method is based on the original data for random tagging and fine-tuning, that is, the original data is modified with random dithering and class tagging. This ensures the randomness of the data, and the linear fine-tuning makes the data more realistic, thus increasing the generalization of the model. To increase the generalizability of our classifier, we may first randomize the perturbation points along the distribution by adding some values from a random distribution. A model trained on these data is more generalizable to those sample data points that are not included in the training set. The smoothed curve is shown in Figure 7a, and the part after performing the data enhancement process appears as discrete points as shown in Figure 7b. The details of the experiments are shown in the Experiments section.



(**a**) (**b**)

Figure 7. (**a**) Before data enhancement; (**b**) After data enhancement.

(1) For body temperature data, 7% to 10% of each dataset is selected for outlier labeling, and the treatment is to add incremental values conforming to a normal distribution (−1 to 2) to the actual data.

(2) For heart rate data, in this paper, 6%~10% of each dataset is selected to mark the outliers, and the processing method is to add increments conforming to a normal distribution (−10~30) to the actual data.
(3) For the ambient temperature and humidity in the cave, 7%~10% of the dataset is selected for outlier labeling.
(4) The relative position information contains a certain meaning for early warning criteria, so there is no need to construct anomalous data.

All data are normalized as a means of converting physical values with wave function characteristics into some relative values that interact with other wave function characteristics by using dimensionless signal processing techniques to reduce the fallout from these quantities [8]. The normalization calculation method chosen in this paper is the method of maximum normalization (min–max normalization, mmn), which can also be referred to as maximum linear normalization or deviation normalization, which is a linear transformation of the original data so that the resulting value can be mapped between [0, 1] or some self-defined interval. The expression of the transformation function is as follows:

$$X' = \frac{X - min}{max - min}$$

where *max* and *min* are the maximum and minimum values of the changed feature dataset, respectively. Some of the data after normalization of the data are shown in Table 7.

Table 7. Data normalization.

Human Body Temperature		Human Heart Rate		Cave Ambient Temperature		Cave Ambient Humidity		Relative Position	
Original Data	Normalization	Original Data	Normalization	Original Data	Normalization	Original Data	Normalization	Original Data	Normalization
36.4	0.27778	62	0.13333	8.9	0.206521	68	0.54906	1	0
36.1	0.592593	110	0.666667	11.3	0.467391	83	0.76702	3	0.2857143
37	0.388889	82	0.355556	9.2	0.239130	62	0.455482	5	0.5714286
37	0.388889	92	0.466667	10	0.326087	68	0.549653	8	1
36.3	0.259259	77	0.3	10.8	0.413043	62	0.452244	4	0.428571
37.1	0.407407	82	0.355556	12.9	0.641304	64	0.482054	4	0.428571
36.3	0.259259	77	0.3	13.4	0.695652	72	0.611903	6	0.714286
36.5	0.296296	73	0.255556	9	0.217391	64	0.482951	2	0.142857
36.5	0.296296	64	0.155556	14.2	0.782608	69	0.559283	6	0.714286
36.9	0.37037	68	0.2	14.8	0.847826	73	0.614246	7	0.857143

5. Experiments

5.1. Performance Evaluation Metrics

In this paper, the performance of the OCSVM-based personnel security status recognition model is evaluated by accuracy, precision, recall, and F-score values [25], and the relevant formulae are shown below.

$$\begin{aligned} \text{accuracy} &= \tfrac{(TP+TN)}{(TP+TN+FN+FP)} \\ \text{precision} &= \tfrac{TP}{TP+FP} \\ \text{recall} &= \tfrac{TP}{TP+FN} \\ f\text{−score} &= 2 * \tfrac{\text{precision} * \text{recall}}{\text{precision} + \text{reall}} \end{aligned}$$

The confusion matrix representation of the identification results is shown in Table 8.

Table 8. Confusion representation of classification results.

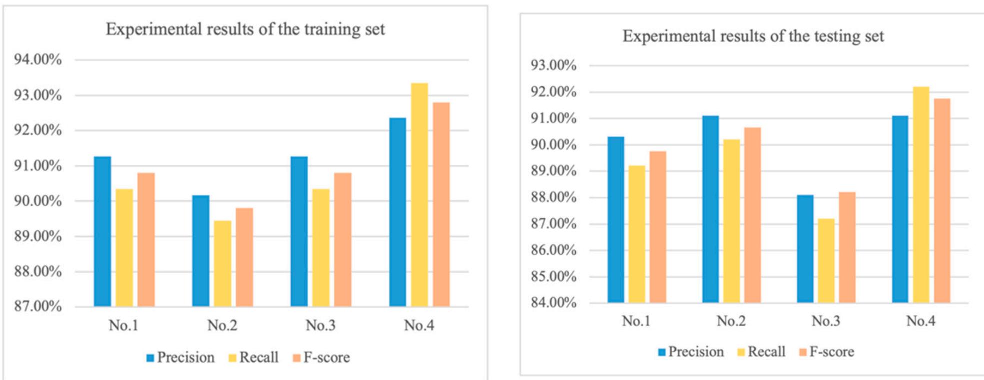
	Actual Positive Class	Actual Negative Class
Judgment Positive	TP	FP
Judgment Negative	FN	TN

5.2. Data Enhancement Experiments

In order to investigate whether the experimental results of the original data after data augmentation have any effect, this subsection conducts an experimental investigation. The main purpose of data augmentation is to make the model training better, i.e., to adopt a suitable data augmentation method when the original data may be too smooth or the amount of data is small. Therefore, the following experiments are designed in this paper to investigate the effect of data augmentation. The control variable method is used to design the validation experiments. First, the datasets 1, 2 and 3 are divided in an 8:2 ratio into the training set and the test set, respectively. Then, the following four experiments are performed to verify the effect of data augmentation.

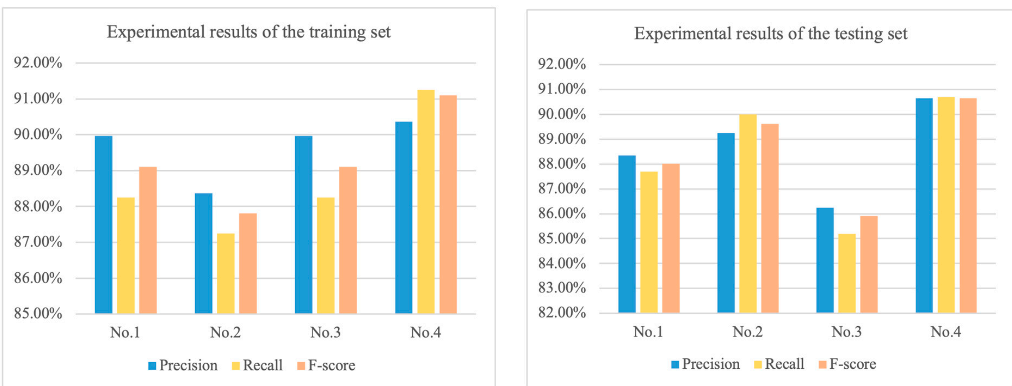
(1) No processing is performed on the three datasets and the OC-SVM model is used for experimental testing.
(2) Data augmentation is performed on the training set part of the three datasets, keeping the test set unchanged, and the OC-SVM model is used for training and testing.
(3) Data augmentation is performed on the test set part of the three datasets, keeping the training set unchanged, and the OC-SVM model is used for training and testing.
(4) The test and training sets of all three datasets are expanded, and the OC-SVM model is used for training and testing.

For the four experiments, which are denoted by No. 1, No. 2, No. 3 and No. 4 in this paper, the precision, recall and F-score are shown in Figures 8–10.



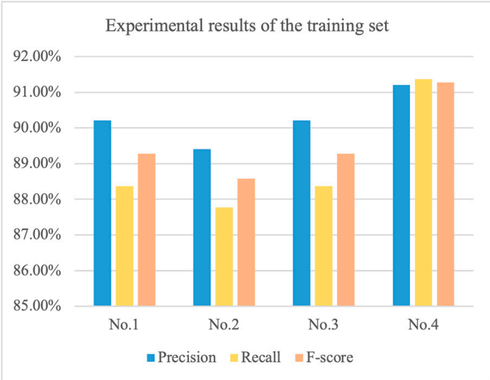
(**a**) (**b**)

Figure 8. Experimental results on dataset 1, where (**a**) shows the training set results, and (**b**) shows the test set results.

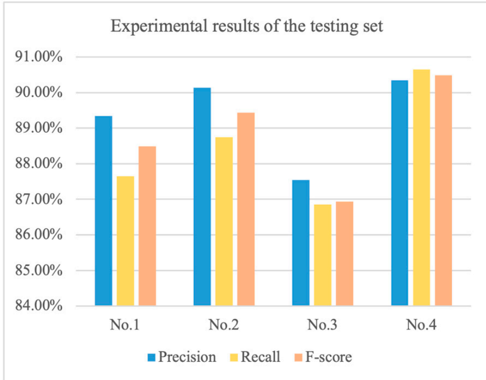


(**a**) (**b**)

Figure 9. Experimental results on dataset 2, where (**a**) shows the training set results and (**b**) shows the test set results.



(**a**)



(**b**)

Figure 10. Experimental results on dataset 3, where (**a**) shows the training set results and (**b**) shows the test set results.

The experimental results are shown above, and it can be seen that the best experimental results occurred in No. 4 in all plots, i.e., the data generalization ability is stronger and the trained model performance is better after data augmentation on the training and test sets, while comparing the results of the training and test sets, the model has overfitting and underfitting. No. 1 has no data augmentation, and the remaining No. 2, 3 and 4 results are analyzed consistently. The models perform around 90% on datasets 1, 2 and 3 with better results after data enhancement. Usually, the evaluation index values of the test set of the model after data augmentation are improved, the evaluation index values of the training set are decreased and the generalization ability of the model is enhanced.

5.3. Comparison Experiments

In this paper, we investigate the effects of three different distributional feature datasets on model performance, distributing four types of data for personnel safety state warning experiments using human feature data, environmental feature data, human–environment data features and human–environment–location information features. The linear kernel function and RBF kernel function were selected for comparison experiments. During the experiments, the human body feature data consisted of body temperature and heart rate, the environmental feature data consisted of internal tunnel ambient temperature and ambient humidity, all of which were two-dimensional data, while the human–environment feature data and human–environment–location feature data were multidimensional data. The ratio of training set to test set is 8:2. The number of support vectors of the models trained by each type of data is counted, and the prediction accuracy of the test set is calculated. The experimental results are shown in Table 9.

After comparing the experimental results of the four types of data with different statistical distribution characteristics in different collection sources, we can find that the performance of the OCSVM model is good on each dataset, and the model accuracy is near 95%. In terms of kernel function selection, the model prediction accuracy of the RBF kernel function is slightly higher than that of the linear kernel function. However, the number of support vectors generated by the RBF kernel function in the models of each dataset is more than that of the linear kernel function, which indicates that the computational complexity of selecting this kernel function is higher. Thus, when the dataset is large, the linear kernel function can be selected for pre-processing and preliminary classification, and then the final state alarm can be determined using the RBF method.

5.4. Ablation Experiments

The data explored in this paper are actual data of tunnel construction personnel, whose anomaly data account for a small proportion of the total data volume. In the study of practical engineering problems, the data proportion directly affects the accuracy and efficiency of the model. Accordingly, the data proportion and balance have a certain degree of influence on the model performance. In order to explore the relationship of this influence, we set the alarm state values in a certain proportion and explore the relationship of (5–50%) proportion of data on the model. The experimental data are dataset 1 (Xinjiang Kuitun River diversion tunnel project (second standard)), the RBF kernel function of the above four experiments is selected as the kernel function, and the accuracy of their test sets are counted at different proportions. The experimental results are shown in Figure 11, where Figure 11a–d represent human characteristic data, environmental characteristic data, human–environment characteristic data, and human–environment–location feature data.

Table 9. Results of personnel safety warning model.

		Human Body Characteristics		**Environmental Characteristics**		**Human–Environmental Characteristics**		**Human Body–Environmental Characteristics–Location Information Characteristics**	
Datasets	Type	Number of SVM	Accuracy	Number of SVM	Accuracy	Number of SVM	Accuracy	Number of SVM	Accuracy
Dataset 1	Linear kernel	59	94%	61	91%	70	94%	76	96%
	RBF kernel	72	98%	79	99%	271	93%	301	91%
Dataset 2	Linear kernel	78	96%	81	95%	62	98%	46	93%
	RBF kernel	69	99%	49	99%	214	94%	242	95%
Dataset 3	Linear kernel	44	97%	86	96%	39	96%	54	95%
	RBF kernel	62	100%	51	99%	217	99%	235	95%

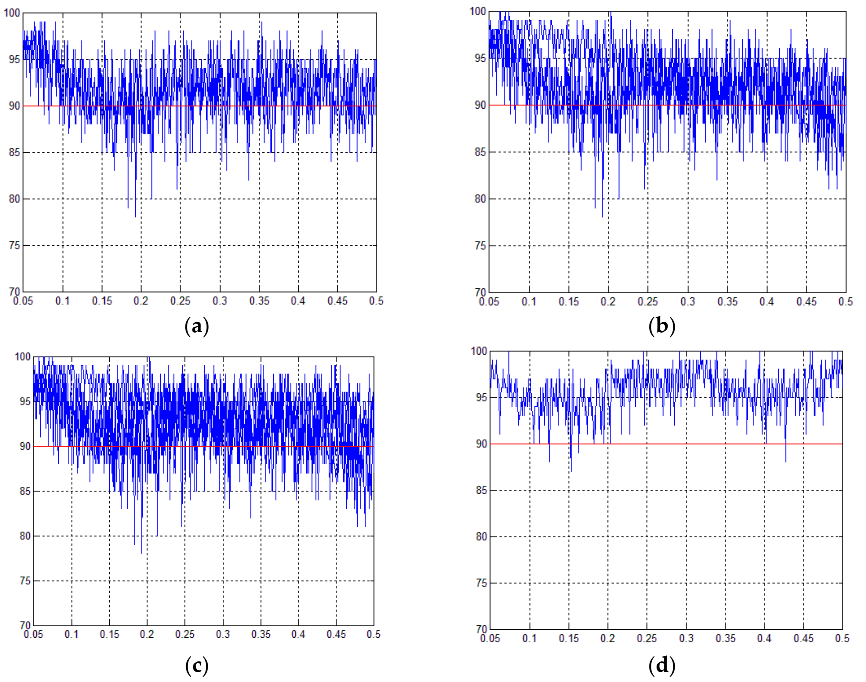


Figure 11. Experimental results. (**a**) represents human characteristic data; (**b**) represents environmental characteristic data; (**c**) represents human–environment characteristic data; (**d**) represents human–environment–location characteristic data. The horizontal axis represents the percentage of abnormal data, and the vertical axis represents the prediction accuracy.

Comparing the experiments with different warning state data ratios, we can see that the prediction accuracy of each type of data is above or below 90% at different warning state ratios, and the classification state is better than other ratios when the data ratio is below 15%. Figure 11d shows that the OCSVM model with the addition of location information can improve the performance of the tunnel staff safety situation early warning, and the accuracy is above 90% for all of them.

Then, in the same experimental conditions, the four cases were compared in a cross-sectional manner, setting the proportion of abnormal data to 15%, selecting the kernel function to choose, and training 100 batches on dataset 1. The experimental results are shown in Figure 12.

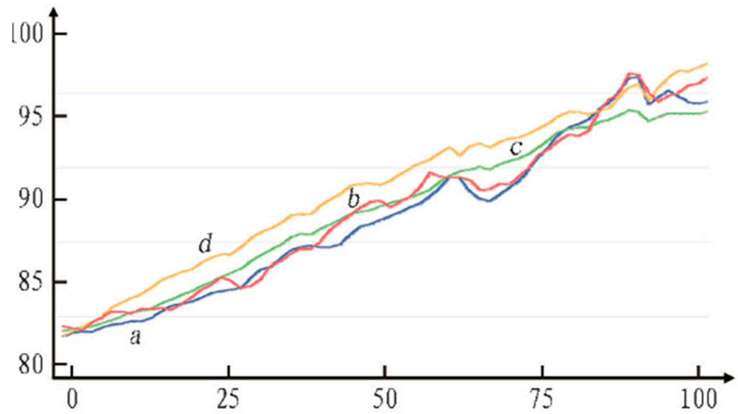


Figure 12. Cross-sectional experimental comparison chart.

The classification effect of the model under all four feature information categories shows an increasing trend, where the classification effect of the model under feature d is ahead of the other three, and it can be seen that the classification performance is better when there is more feature information under the same experimental conditions.

Finally, we experimentally compare the algorithm proposed in this paper with the KNN algorithm, a supervised learning algorithm, which is one of the most classical machine learning algorithms for classification and regression. It is based on the idea that if the majority of the K most similar (i.e., most adjacent) samples in the feature space belong to a class, then that sample also belongs to that class. That is, the method decides the category to which the sample to be classified belongs based only on the category of the nearest one or more neighboring samples in the decision classification. The experimental results are shown in Table 10.

Table 10. Comparison of experimental results using OCSVM model and KNN model.

NO.		Precision	Recall	F-Score
Dataset 1	OCSVM	92.37%	93.34%	92.80%
	KNN	82.12%	82.21%	81.75%
Dataset 2	OCSVM	90.37%	91.25%	91.10%
	KNN	79.14%	80.32%	80.63%
Dataset 3	OCSVM	91.21%	91.37%	91.28%
	KNN	80.21%	80.55%	80.42%

As can be seen from the data in the table, the detection effect of OCSVM is improved compared with the KNN algorithm, with an average improvement of about 10% for each index. Therefore, the OCSVM method is more accurate in detecting the safety condition of tunnel personnel and improves the safety of tunnel personnel. In conclusion, we believe that we have provided an efficient and concise machine learning algorithm that can achieve better identification and classification despite the large deviation of positive and negative samples in the dataset (i.e., the number of positive samples may be much larger than the number of negative samples), and can be successfully applied to the field of tunnel personnel safety status detection.

6. Conclusions

In this paper, an OCSVM-based method for tunnel construction personnel's safety status detection is proposed, and three sets of data on the physical signs and environmental characteristics of tunnel construction safety personnel are collected. Several variables in each dataset were selected separately to build the corresponding OCSVM safety warning model. The model was applied to detect the personnel safety information of the test set. By analyzing the detection results of the data with different feature distributions, we can obtain a model accuracy above 90%, which verifies that the OCSVM algorithm has good scientific generalization ability and scalability for early warning based on the tunnel personnel safety information. We conducted ablation experiments with different data ratios in different warning states, i.e., data experiments with different balance degrees, which showed better performance and proved the good scientific generalization ability and scalability of the model. Moreover, this paper compares the OCSVM model with the KNN algorithm. In the three datasets, the evaluation indexes precision, recall and F-score are all nearly 10% higher for the OCSVM model than the KNN algorithm, which proves that the OCSVM algorithm is more suitable for tunnel personnel safety detection and is conducive to guaranteeing the safety of tunnel construction personnel.

For future research on the tunnel personnel safety early warning problem, the next step could be to expand the data features, multi-dimensional classification and prediction of the OCSVM algorithm. The research direction is to expand the amount of data and data dimensions for experiments, combine them with deep learning models, complete data unsupervised training, and compare with a variety of machine learning prediction and classification methods to better serve the tunnel and civil construction.

Author Contributions: Conceptualization, G.H.; methodology, G.H.; software, G.H.; validation, G.H., J.C. and L.L.; formal analysis, J.C.; investigation, L.L.; resources, L.L.; data curation, L.L.; writing—original draft preparation, L.L.; writing—review and editing, G.H.; visualization, G.H.; supervision, G.H.; project administration, G.H.; funding acquisition, G.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Materials used for experiments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sun, Z.; Qian, T.; Ren, Y.; Chu, C.; Zhou, Z. Study on key technologies and application of engineering big data management platform of tunnel boring machine. *Tunn. Constr.* **2020**, *40*, 783–792.
2. Chen, Y. Application of Intelligent algorithm based on genetic algorithm and extreme learning machine to deformation prediction of foundation pit. *Tunn. Constr.* **2018**, *38*, 941–947.
3. Cao, H.; Yang, L.; Lin, Y.; Qu, C. Aero-engine fault detection based on one-class support vector machine. *J. Civ. Aviat. Univ. China* **2013**, *31*, 26–28.
4. Cheng, W.; Yan, W.; Tingfei, Y. Improving the detection accuracy of wake targets by improved one-class support vector machine. *Acta Armamentarii* **2020**, *41*, 1887–1893.
5. Gao, Y. Fault detection of gas insulated switchgear based on one class support vector machine. *Electrotech. Electr.* **2020**, *2020*, 52–55.
6. Zheng, H. Research on network traffic anomaly detection in distribution communication network based on OCSVM. *Inf. Commun.* **2018**, *10*, 97–98.

7. Qin, J.; Cao, X.; Jiao, L. Detection and tracking of hard hat wearing based on deep learning. *Comput. Mod.* **2020**, *6*, 1–6.
8. Long, C.; Yang, M.; Li, J.; Li, W.; Xiong, Z. Experimental research on safety early warning system of live working based on human physical signs. *Hunan Electr. Power* **2020**, *40*, 55–60.
9. Yan, F.; Dong, L.; Wang, B.; Ge, J.; Wang, B. Using risk meshing to improve three-dimensional risk assessment of chemical industry. *Process Saf. Environ. Prot.* **2022**, *168*, 1166–1178. [CrossRef]
10. Yu, Z.; Shao, Z.; Hu, J.; Bu, L.; Bian, H. Cloud model-based safety of university laboratory management standardization research. *Lab. Res. Explor.* **2020**, *39*, 300–304.
11. Peng, T.; Li, C.; Zhou, X.B. Application of machine learning to laboratory safety management assessment. *Saf. Sci.* **2019**, *120*, 263–267. [CrossRef]
12. Zhang, Y.J.; Lin, Y.; Qian, J.; Zhou, J. DHGF algorithm-based laboratory safety management in higher education institutions room safety management evaluation. *Lab. Res. Explor.* **2016**, *35*, 300–303.
13. Tang, J.W.; Dong, B.; Kong, J.G. Fire safety evaluation of university laboratories based on cloud object metamodel safety evaluation research. *Exp. Technol. Manag.* **2021**, *38*, 270–274.
14. Ren, Y. Application of fuzzy comprehensive evaluation in laboratory safety management evaluation. *China China Saf. Prod. Sci. Technol.* **2015**, *11*, 186–190.
15. Schölkopf, B.; Platt, J.C.; Shawe-Taylor, J.; Smola, A.J.; Williamson, R.C. Estimating support of a high-dimensional distribution. *Neural Comput.* **2001**, *13*, 1443–1471. [CrossRef]
16. Li, C.; Wang, B.; Tian, J.; Guo, R. Anomaly detection method for UAV sensor data based on LSTM-OCSVM. *J. Chin. Comput. Syst.* **2021**, *42*, 700–705.
17. Mao, Q.; Liang, Z.; Liu, D.; Hu, Y.; Li, G.; Fang, X. Fault detection of temperature sensor in chiller based on one-class support vector machine algorithm. *J. Refrig.* **2019**, *40*, 130–142.
18. Zhou, Y.; Li, J.; Song, Q. A web page information extraction method based on SVM and Text density features. *Comput. Appl. Softw.* **2019**, *36*, 251–261.
19. Yuan, Y.; Zhong, P.; An, Z.; Wang, H. A new multi-class support vector machine algorithm. *Math. Pract. Theory* **2011**, *41*, 94–100.
20. Li, X.; Li, R.; Zeng, Y. Classification of support vector machines based on improved kernel function on time series data. *J. Xinyang Agric. For. Univ.* **2021**, *31*, 121–126.
21. Ye, Q.; Zhao, C.; Ye, N.; Chen, X. Localized twin SVM via convex minimization. *Neurocomputing* **2011**, *74*, 580–587. [CrossRef]
22. Lu, S.; Tong, L.; Zhu, C. Least squares projection twin support vector machine with universum. *J. Hebei Univ.* **2016**, *36*, 94–99.
23. Wang, X.; Liu, W.; Zhang, M. Investigation about the effects of temperature and humidity on personnel in Qinling railway tunnel. *J. North. Jiaotong Univ.* **1999**, *23*, 65–69.
24. Tang, Z.; Xu, Q. Rockburst prediction based on nine machine learning algorithms. *Chin. J. Rock Mech. Eng.* **2020**, *39*, 773–781.
25. Wang, J. Study on control of hot and humid environment in TBM driving of Zhongtianshan extra-long Tunnel. *Tunn. Constr.* **2014**, *34*, 212–215.