Article

A Machine Learning-Based Risk Assessment Study of Unsafe Behavior of Workers in Nuclear Power Plants Under Construction

Xueqiang Shan [1], Weibo Yang [1], Xia Liu [2,*], Kai Yu [3,*] and Hui Cui [3]

[1] General Manager Department, Zhangzhou Project Department, China Nuclear Power Engineering Co., Ltd., Beijing 100840, China; safety769434838@foxmail.com (X.S.); safety1246441565@foxmail.com (W.Y.)

[2] Project Control Department, Zhangzhou Project Department, China Nuclear Power Engineering Co., Ltd., Beijing 100840, China

[3] College of Safety and Environmental Engineering, Shandong University of Science and Technology, Qingdao 266590, China; safety1838567469@foxmail.com

[*] Correspondence: lxzzhdlx@163.com (X.L.); ykxfkd@sdust.edu.cn (K.Y.);
Tel.: +86-17660923975 (X.L.); +86-13969849358 (K.Y.)

Abstract: Most accidents during the construction of nuclear power plants are caused by human unsafe behavior. How to scientifically determine the risk management priority of human unsafe behaviors is the basis for effectively preventing accidents in under-construction nuclear power plants. Although employees are adopted for control in under-construction nuclear power plants, the records of unsafe behaviors are mostly recorded by inspectors, and the records of behaviors may have missing values. To overcome the above problems, this paper applies machine learning algorithms to construct an employee behavioral risk assessment model. Firstly, by analyzing the influencing factors of unsafe behaviors, the assessment indexes are proposed, then the Random Forest algorithm is used to obtain the characteristic importance of the proposed indexes and exclude those with smaller characteristic importance. Finally, the harmony search (HS) algorithm is used to optimize the back propagation (BP) neural network to construct an assessment model and compare with the BP evaluation model. The results show that the HS-BP model is more accurate and efficient. The results show that the method can comprehensively and effectively analyze workers' unsafe behaviors, and the BP neural network is optimized to construct the assessment model using the Harmonic Search algorithm, which is more accurate than the original model. The use of the machine learning method to assess workers' behaviors can objectively output the risk level and overcome the one-sidedness and subjectivity of the traditional expert evaluation method.

Keywords: risk assessment; nuclear power; under construction; machine learning; assessment

1. Introduction

With the rapid development of economic construction, China's nuclear power development strategy has been adjusted from moderate development to active development. At present, a number of nuclear power projects have been approved by the state and entered the stage of full-scale construction. As an industry with extremely high safety requirements, nuclear power enterprises must do a good job of their own safety production

in accordance with relevant laws and norms. The construction sector is responsible for the actual construction of nuclear power plants, which is the base scenario for the occurrence of various types of human behavior. Human unsafe behavior also occurs mostly during this construction process. As for the control of employees involved in the construction of nuclear power plants, and despite the current control measures, there is still the problem of unsafe behaviors being recorded by inspectors, leading to the possibility of missing values. In recent decades, scholars at home and abroad have conducted plenty of pioneering and fruitful research on safety behavior, as well as accumulated rich research results and research experience.

After the nuclear power accidents at Three Mile Island and Chernobyl, the importance of human behavior to the safe operation of nuclear power plants has drawn widespread attention. Human Factors Engineering (HFE) has become an important factor that must be taken into consideration in the design of nuclear power plants [1,2]. The design process of nuclear power plants involves twelve elements of human factors engineering, and Human Reliability Analysis (HRA) is an important part among them [3]. Generally, the HRA method includes two parts, namely, qualitative analysis and quantitative analysis. However, in the application process, in the past, more attention was paid to the calculation results of quantitative analysis, while qualitative analysis was often overlooked or merely regarded as a process of quantitative analysis. In fact, qualitative analysis and quantitative analysis play equally important roles in HRA. To emphasize the importance of qualitative analysis, in 2012, O'Hara et al. [4] from the U.S. Nuclear Regulatory Commission (NRC) renamed HRA as the "Treatment of Significant Human Actions" in the third edition of the "Human Factors Engineering Review Program" (NUREG-0711) and emphasized that detailed analysis should be conducted on safety-related human behaviors during the design process.

As a complex large-scale man–machine system, nuclear power plants involve a large number of different types of personnel behaviors [5]. At the beginning of the design of nuclear power plants, designers need to analyze and process these personnel behaviors [6] to ensure the smooth completion of the subsequent design work. The personnel behaviors here refer to the behaviors of operators in nuclear power plants [7], because most behaviors in nuclear power plants are executed by operators. Different personnel behaviors have different impacts on the safe operation of nuclear power plants. Identifying, screening, and classifying personnel behaviors and then carrying out corresponding processing can not only effectively improve the efficiency of design work but also provide a priority ranking for personnel behavior processing to optimize man–machine interaction tasks and interfaces. More importantly, it can ensure that important personnel behaviors related to the safe operation of nuclear power plants are not overlooked.

In 2002, Higgins et al. [8] first proposed and established a personnel behavior screening method, laying the foundation for personnel behavior screening. This method is based on the Probabilistic Safety Assessment (PSA) of nuclear power plants, so it is called the probabilistic screening method. The probabilistic screening method has been widely used. It completes the screening by evaluating the Risk Achievement Worth (RAW) and the Fussell-Vesely (FV) of personnel behaviors [7–9]. The advantages of the probabilistic screening method lie in its relatively simple screening and good traceability; its disadvantage is that it relies on the results of PSA analysis and inherits the uncertainties of PSA [10]. Therefore, if only the probabilistic screening method is used to identify and screen personnel behaviors, the screening results will not meet the requirement of being fully conservative.

In 2007, Zhao Jun et al. [11] expounded the complementary relationship between the PSA method and the deterministic analysis method in the process of equipment classification, solving the problem that using a single method in the process of equipment risk

classification cannot ensure full conservatism. Although equipment classification and personnel behavior screening have different implementation objects, they are both derived from the same theoretical basis.

Human unsafe behavior is the main cause of safety accidents [12]. The main influences on unsafe behavior include organizational and social factors, managerial safety, work safety tension, safety climate and personal safety awareness [13]. As research has progressed, it has been found that factors contributing to unsafe behaviors exist in many areas. For example, the behavior of leaders and management policies will have an impact on the behavior of employees [14]. The lack of safety culture will lead to unsafe behaviors of employees [15], and individual perception, environmental support, the experience of organizational management systems, and individual psychological stress will also lead to unsafe behaviors [16].

Based on the individual, production operations, organizational management, four dimensions of the social environment, the construction of workers' unsafe behaviors' influencing factors, and the ISM (Interpretive Structural Model), Ye Gui et al. conducted a systematic analysis of the role of factors in various dimensions of the effect of the individual factors for the surface factors and the social environment for the conclusion of the deep factors [17]. Based on the perspective of the organizational climate, Cheng Jialei and Qi Shenjun constructed a relationship model between the organizational climate, individual factors (safety attitude, unsafe motivation, safety ability), and unsafe behaviors [18]. How to accurately assess the risk of safety behaviors of field operators is an urgent problem. In risk practice, some scholars directly use accident potential data to calculate risk parameters, and the accuracy of risk magnitude is largely affected by the quality of the data [19]. Younggi Hong and Jaeho Cho's study focuses on improving pre-emptive risk identification and safety checks to prevent workplace accidents [20]. Human factors are the main part of occupational risks, and reducing occupational risks is one of the important methods of OSH management [21]. Deng et al. determined that, in all industries, safety behaviors play a vital role in preventing accidents and injuries [22].

Safety behavioral science originates from the behavioral sciences, is a cross-discipline of behavioral sciences and safety sciences, and European and American scholars started the research in this area earlier [22]. British scholars Gene Earnest and Jim Palmer in 1979, for the first time in the name of behavioral safety management (behavior-based safety, BBS), put forward the study of safety behavior [23]. At the end of the 1970s, domestic scholars began the study of safety behavioral science. Since the 21st century, with the rapid development of safety science, the field has achieved many results. At present, the more popular basic principles of safety behavior science include the following: Maslow's [24] hierarchy of needs theory; Albert Bandura's [25] social learning theory; Victor H. Vroom's [26] expectancy theory; and Icek Ajzen's [27] theory of planned behavior. This method is used to prevent and correct unsafe behaviors by observing, measuring, feeding back, and reinforcing the behaviors of the observed person. Foreign countries have an early start in safe behavior modification research and have accumulated a wealth of experience and results. The early formation of more famous theories are Greenwood M's accident tendency theory [28], Heinrich H W's accident causal chain reaction [29], Surry J's Thurley model [30], Reason J's human-caused accident causal model [31] and so on, based on the research that scholars conducted on the influence of human unsafe behaviors, the formation of the mechanism of in-depth research, and the factors leading to unsafe behaviors. It is found that the factors leading to unsafe behaviors are mainly individual, environmental, and organizational management factors at three levels. The research and application of safe behavior modification in Europe, the United States, and other industrially developed countries are more mature, have been successfully applied to a number of industries, and have achieved significant results. Although domestic research started later,

it has also made significant progress in recent years. Cao Qingren and others tried to explain unsafe behaviors from the perspective of cognitive psychology and put forward the 'know-can-do' model of unsafe behaviors [32].

Domestic scholars have conducted a large number of empirical studies on the basis of foreign advanced experience, combined with national conditions and industry characteristics, and put forward many safe behavior modification methods and techniques applicable to the domestic environment.

In order to conduct a more accurate study of the risk of unsafe behavior of workers in nuclear power plants under construction in Zhangzhou, Fujian, this paper introduces the BP neural network.

The BP (back propagation) neural network [33] is a multi-layer feed-forward neural network trained according to the error back propagation algorithm, which is the most widely used neural network. The BP neural network can make the output error reach the preset value within the range through the back propagation of error signals, and the error signal back propagation process can be adjusted according to the error gradient descent method to achieve the output close to the expected value. As a multi-layer network model, the number of hidden layers can be single layer or multi-layer in the BP neural network, but the more hidden layers are used, the more complex the calculation of the backward propagation process of the BP network error, and it is easier to fall into the local optimal situation [34]. The BP neural network is currently the most widely used artificial neural network model. It is a black-box model built to simulate the process of information processing by the human brain. Since it is not necessary to use sophisticated mathematical equations to obtain more accurate predictive values, it is widely used in the prediction of various nonlinear relationships [35].The BP neural network algorithm (back propagation, BP) is a mathematical model that can deal with complex nonlinear problems [36–38].

The BP neural network has the advantages of strong nonlinear ability, fault tolerance, and rigorous derivation process; however, the algorithm still has the defects to easily fall into local minima and slow convergence speed. In order to solve these problems, in this study, the HS-BP assessment model is established by combining the harmonic search algorithm with strong global search capability with the BP neural network, optimizing the initial weights and thresholds of the BP neural network. The results of this study can accurately assess the risk level of employee behavior, identify the people who need priority control, and provide the basis and theoretical support for the safety management of nuclear power plants under construction, which has strong theoretical value and practical significance.

2. Materials and Methods

2.1. Advantages of Machine Learning Technology

In the operation management system of nuclear power plants, an accurate assessment of the risk of unsafe employee behavior is crucial to ensure the safe and stable operation of nuclear power plants. Current machine learning models used for this assessment differ significantly in multiple key dimensions compared to traditional existing models.

In terms of model construction, the traditional model is based on traditional experience and simple statistics, with limited factor identification, and its simple linear or rule-based system architecture is weak in capturing complex nonlinear relationships, whereas the machine learning model comprehensively analyzes many influencing factors with the help of advanced algorithms such as random forests, filters the key features, and optimizes the model with BP neural networks and HS algorithms, which avoids falling into local optimal solutions during the training of BP neural networks and thus more accurately captures the complexity of safety behavior in a more precise way. The model is based on the BP neural network and is optimized with the HS algorithm to avoid local

optimal solutions during BP neural network training, thus capturing the potential patterns and relationships behind complex data more accurately. In terms of performance, the traditional model is not capable of handling complex factor relationships, and its inability to accurately grasp these relationships leads to poor accuracy of the assessment results, and it is extremely sensitive to data fluctuations and susceptible to stability, making it difficult to continue to provide a reliable assessment; the machine learning model, through the well-designed feature screening process and advanced algorithmic optimization strategy, achieves a significant increase in accuracy and stability and maintains a high level of accuracy under data variations in different time periods and working conditions. Machine learning models, on the other hand, have achieved significant improvements in accuracy and stability through a carefully designed feature screening process and advanced algorithm optimization strategies and are able to maintain a high level of assessment under different time periods and working conditions. In data processing, traditional models do not make full use of data, are weak in dealing with missing values, and rely on complex and scientifically insufficient manual filling or correction methods, which are time-consuming, labor-intensive, and introduce bias, thus affecting the accuracy of the assessment. Machine learning models, by virtue of their unique algorithmic advantages, can effectively deal with multi-dimensional data and are robust to missing values, like the random forest algorithm, which can, to a certain extent, ignore the impact of missing values in calculating the importance of features, and the BP neural network, which can effectively deal with the data from the perspective of the characteristics of the model. The BP neural network can learn from the overall data features to reduce the interference of missing values on the model performance. In terms of application adaptability, the traditional model has poor scene adaptability and limited real-time and dynamic assessment capability, which makes it difficult to adapt to the complex and changing operating environment of the nuclear power plant and unable to accurately carry out risk assessments in a timely manner. The machine learning model, by virtue of its strong learning capability and high flexibility, can better adapt to the dynamic changes in different scenes. It is especially suitable for a real-time risk early warning, and it can be used to quickly analyze and provide early warnings based on real-time data when there is an unexpected situation or parameter change in the operation of the nuclear power plant. When unexpected conditions or parameter changes occur in the operation of the nuclear power plant, it can be quickly analyzed and warned based on real-time data.

Based on the above discussion, we can learn that machine learning models have significant advantages in the risk assessment of unsafe behaviors of nuclear power plant workers, which strongly supports the improvement of the safety management level of nuclear power plants. Especially in dealing with the risk assessment of unsafe behaviors of employees in nuclear power plants under construction, machine learning technology has the following outstanding advantages: Firstly, it can automatically deal with the missing value data by virtue of a unique algorithmic mechanism, overcoming the lack of a scientific nature of the traditional manual filling method and the problem of easy introduction of bias. And it optimizes the initial weights and thresholds by constructing the HS-BP model, combining with the harmony search algorithm and the back propagation neural network, avoiding the local optimums, and fitting the complex nonlinear relationships better, which significantly improve the accuracy and efficiency of the model, provide prompt assessment results when dealing with a large amount of data, and offer powerful support for risk management. It can provide timely and powerful support for risk management, fully demonstrating the powerful ability of machine learning technology to optimize model performance.The advantages of machine learning are shown in Table 1.

Table 1. Advantages of machine learning over traditional methods.

Comparison Project	Traditional BP Models	HS—BP Model
Model optimization approach	The use of the back propagation algorithm based on a gradient descent to adjust the weights and thresholds, easy to fall into the local optimal solution, limiting the ability of model generalization.	Strong global search capability to jump out of local optimal solutions.
Convergence speed and accuracy	Dealing with complex problems, long training time, limited prediction accuracy due to the tendency to fall into local optima.	Fits data more accurately, learns nonlinear relationships more efficiently when dealing with complex problems, better accuracy and efficiency.
Model stability	Sensitive to initial weights and threshold settings, different initial values can lead to large differences in training results and inconsistent performance when run multiple times or applied to different datasets.	Higher stability and reliability, more consistent performance in different scenarios.

2.2. Indicators to Be Selected

The accident chain theory clarifies the various causes of casualties and the relationship between accidents and that each casualty is not an isolated event but is the result of a series of events occurring one after the other, even though the injury may occur suddenly at a certain moment. The interaction and correlation between many factors makes it impossible to fully grasp the direct and indirect causes of unsafe behaviors of workers in nuclear power plants under construction. However, through long-term research and work experience, we have gained knowledge of the factors affecting employee behavior, which makes it possible to find the important influences on employee behavior.

The human factor is an important cause of unsafe behaviors, and personal characteristics such as gender, age, education, and years of experience have a significant impact on the behavior of employees. Accidents are caused by imperfect safety warning signs; incomplete or ineffective safety protection equipment and other problems occur from time to time. The temperature, noise, and harmful gas concentration of the construction environment will have a certain impact on the behavior of construction workers when considered in conjunction with human sensory perception and the characteristics of the site of production activities. In terms of management, the influence of safety supervision and inspection, work arrangement, safety education and training, and emergency management on safety behavior is mainly considered in construction management.

The selection of indicators in this paper is based on the theory of accident chaining. Through long-term research and work experience, we learned the main reasons that lead to the occurrence of unsafe behaviors of employees in nuclear power plants, as a way to obtain the relevant factors affecting the behavior of employees. Since the purpose of this paper is to use machine learning methods to scientifically and reasonably assess the unsafe behaviors of workers in nuclear power plants under construction, the indicators that we selected need a large amount of actual data as a basis. Considering that the actual data of some indicators are not easy to obtain or difficult to quantify, and combining with the existing data resources, the indicators initially selected are shown in Table 2.

Table 2. Behavioral evaluation-related indicators.

Dimension	Norm
indicator personal factors	age, length of service, attendance, psychological condition, knowledge and skills, safety awareness
environmental factor	noise, temperature, welding fumes, construction dust
organizational factors	safety supervision and inspection, work organization, safety education, emergency management
machinery and equipment	safety warnings, equipment reliability, protective gears

2.3. Methodologies

Machine learning refers to a methodology that can automatically detect patterns in data, and these methods can be used to develop predictive models and aid decision making under conditions of uncertainty. Machine learning includes research on decision trees, random forests, and artificial neural networks. With the rapid development of China's nuclear power business, the ability to obtain information about unsafe worker behavior has been greatly improved, but the lack of information processing capability has resulted in the information not being used effectively. Therefore, in order to improve the safety management of nuclear power plants under construction, it is necessary to apply machine learning to safety management.

When dealing with the risk assessment of employee unsafe behaviors in nuclear power plants under construction, the random forest algorithm can calculate the importance of feature variables to achieve feature selection and dimensionality reduction, avoiding overfitting, and capturing complex data relationships to enhance model stability; the BP algorithm has a powerful nonlinear mapping capability, is widely used and mature, and has trainability and flexibility; the harmony search algorithm can make up for the defects of BP algorithm, which is easy to fall into the local optimization. Its global optimization and heuristic search advantages complement the BP algorithm, and the combination of these three algorithms provides powerful support for the effective assessment of the risk of unsafe behaviors of workers.

2.3.1. Random Forest

Random forest is a general machine learning algorithm that can handle classification and regression. It can also complete data dimensionality reduction, outlier processing, and data analysis. The main computational steps of random forest designed in this paper are as follows:

(1) The variable importance score is denoted by VIM, the Gini index is denoted by *GI*, and the Gini index score VIM is calculated for each feature X_j by taking m features $X_1, X_2, \ldots, X_m$. The Gini index is calculated as follows:

$$GI_m = \sum_{K=1}^{|K|} \sum_{k' \neq k} P_{mk} P_{mk} = 1 - \sum_{K=1}^{|K|} p_{mk}^2 \qquad (1)$$

where *k* indicates that there are *K* categories; P_{mk} indicates the proportion of *k* in the category column in node m.

(2) The importance of feature X_j in node m, i.e., the amount of change in the Gini index before and after the branching of node m.

$$VIM_{jm}^{(Gini)} = GI_m - GI_l - GI_r \qquad (2)$$

where GI_l and GI_r denote the Gini index of the two new nodes after scoring, respectively.

(3) If the node where feature X_j appears in decision tree i is in set M, then the importance of X_j in the ith tree is

$$VIM_{ij}^{(\text{Gini})} = \sum_{m \in M} VIM_{jm}^{(Gini)} \tag{3}$$

Suppose there are n trees in the RF.

$$VIM_{j}^{(\text{Gini})} = \sum_{n=1}^{n} VIM_{ij}^{(Gini)} \tag{4}$$

(4) Simply normalize all the importance scores obtained.

2.3.2. Differences Between HS-BP and Traditional BP Models

The HS-BP model was developed to overcome the shortcomings of traditional methods and simple BP models in dealing with the risk assessment of unsafe behaviors of employees in nuclear power plants under construction, to achieve a more scientific, accurate, and efficient risk assessment and to provide strong support for the prevention of accidents.

The HS-BP model differs from the traditional BP model in many aspects.

(1) Model Optimization

The BP neural network adjusts the weights and thresholds of the network through the back propagation algorithm to minimize the error between the predicted output and the actual output. However, the traditional BP algorithm has some defects, such as being easy to fall into the local optimal solution. This is because its optimization method is based on the gradient descent. In the presence of multiple local minima on the error surface, once the algorithm converges to a certain local minima, it is difficult to jump out of the algorithm to search for the global optimal solution, resulting in the limited ability of the model to generalize.

The HS-BP model introduced the harmonic search (HS) algorithm to optimize the BP neural network on the basis of the traditional BP model. The harmony search algorithm simulated the process of music harmony creation and searched for the optimal solution by continuously adjusting the parameters of the harmony (corresponding to the weights and thresholds of the neural network). The algorithm has a strong global search ability, which can help the BP neural network to jump out of the local optimal solution, and is more likely to find the global optimal combination of weights and thresholds, so as to improve the performance of the model.

(2) Convergence Speed and Accuracy

The convergence speed of traditional BP models is relatively slow, especially when dealing with complex problems or large-scale datasets, which requires longer training time to achieve better performance. Moreover, due to the tendency to fall into local optimality, the accuracy of the final prediction may be limited, and the potential patterns in the data cannot be fully explored.

The HS−BP model takes advantage of the global search property of the harmonic search algorithm to find better weights and thresholds in a shorter time, which accelerates the convergence speed of the model. Meanwhile, by avoiding the problem of a local optimal solution, the HS−BP model can fit the data more accurately and improve the accuracy of prediction. When dealing with complex problems such as the risk assessment of unsafe behaviors of employees in nuclear power plants under construction, the HS−BP model is able to learn complex nonlinear relationships from a large amount of data more efficiently, thus outperforming the traditional BP model in terms of both accuracy and efficiency.

(3) Model Stability

Since the traditional BP model is sensitive to the initial weights and thresholds, different initial values may lead to large differences in model training results and poor stability. This makes the performance of the model inconsistent when the model is run multiple times or applied to different datasets.

The HS—BP model reduced the sensitivity of the model to initial values by optimizing the weights and thresholds with a harmonic search algorithm. The harmonic search algorithm searched for the optimal solution globally, which makes the model converge to a better result under different initial conditions, improves the stability and reliability of the model, and makes its performance more consistent in different scenarios.

2.3.3. Establishment of the HS-BP Model

The BP neural network usually consists of an input layer, output layer, and hidden layer connected by a set of weight factors. The BP neural network calculated the final network error according to the forward operation of the input data, transmitted the error in the opposite direction, and adjusted the weights and thresholds of the corresponding layers according to certain rule mechanisms when it passed through different layers. After training with a large number of data samples, the final algorithmic model that can complete complex nonlinear mapping was constructed.

The harmony search algorithm analogized instrument i (i = 1, 2, …, m) to the ith design variable in the optimization problem, the harmony R_j (j = 1, 2, …, M) of each instrument's voice tone was equivalent to the jth solution vector of the optimization problem, and the evaluation analogized to the objective function. The algorithm first generated M initial solutions (harmonies) into the HM (harmony memory), searched for new solutions within the HM with probability HMCR, and searched in the domain of possible values of the variables outside the HM with probability 1-HMCR. The algorithm then generated a local perturbation to the new solution with probability PAR. It determined whether the new solution objective function value is better than the worst solution within the HM and if so, replaces it and then iterates continuously until a predetermined number of iterations, T_{max}, is reached.

(1) Initialize algorithm parameters: Set the corresponding parameters in the algorithm procedure, HMS represents the memory bank inventory, HMCR represents the corresponding probability value, and PAR represents the corresponding audio frequency. Let h represent the step length, and let N be the corresponding number of values.

(2) Initialization of the harmonic memory bank: After random transformation, create a new general memory bank containing harmonies.

(3) Create new harmony: The creation of a new harmony is mainly performed by analogy in the memory bank and setting a new tone. Learning or randomly creating a tone through the harmonic memory bank produces a component.

(4) Updating the harmonic memory bank.

(5) Judging the optimal solution: The value derived above determines whether it can be the optimal solution. If so, the output is performed. If it is not satisfied that it becomes the optimal solution, rounding, repeat the two steps 3 and 4 until the optimal solution is obtained, ending the algorithm.

In the HS algorithm, in order to determine its coding length, it is necessary to determine its corresponding neural network first, and, through the study of its neural network, the target length is obtained. For the HS algorithm, it is necessary to first derive its corresponding threshold value and weight value through the corresponding functional relationship substitution, select the target value in the memory bank, assign the value to it,

and derive the optimal solution by using the functional relationship equation, and, after that, improve it to a certain extent by using the BP neural network model, so that the optimal solution is projected out on the neural network. The steps to optimize the BP neural network by the harmonic search algorithm are as follows:

(1) Construct the BP neural network model: Determine the input, hidden, and output layers of the BP neural network according to the corresponding sample features.

(2) Coding: For the HS algorithm, the coding of its algorithm is usually real number coding in academic terms, and, in general, the parameters w_{ij} and θ_j are coded so that dimension m can be obtained by the algorithm, and the extracted harmonies will have the corresponding weights and thresholds.

(3) Determine the fitness function of the HS algorithm: For the group in the memory bank, using the fitness function to calculate it can show the strengths and weaknesses of this group in some sense and to some extent.

(4) HS optimization of BP algorithm: According to the calculation process of the HS algorithm, the initial weights, thresholds, and global extremes of the BP neural network structure are subjected to the optimal and acoustic search as well as the optimality seeking calculation.

(5) Determination of end conditions: For the conclusion of the output, it is necessary to compare it. If it meets the requirements, output the result, and the result is the optimal solution. If it does not meet the requirements, discard the result, which cannot be used as the optimal solution, and repeat the HS algorithm until the output of the result meets the conditions, and the optimal solution is obtained.

(6) Output the global optimal solution.

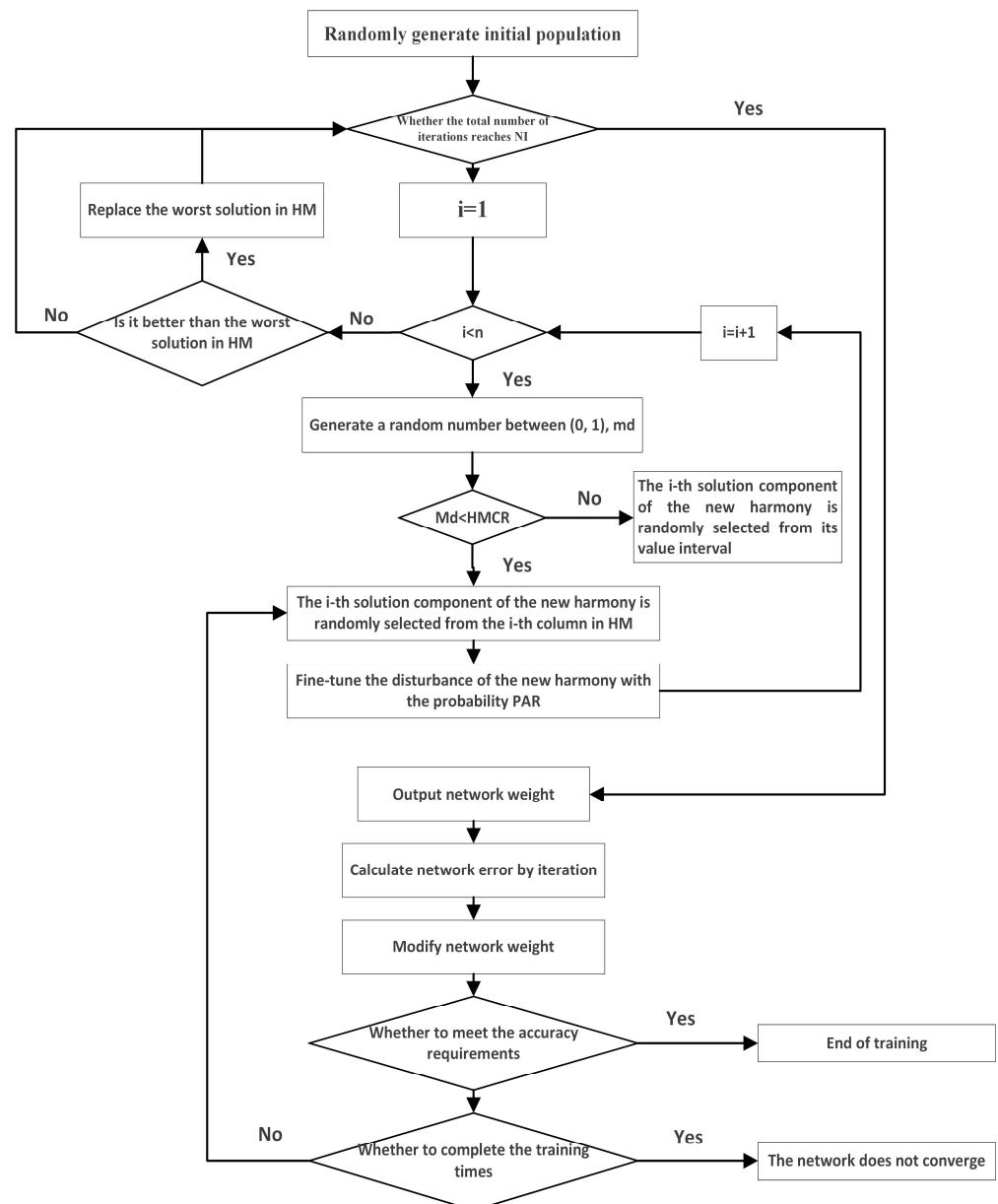The specific flow of the improved algorithm is shown in Figure 1:

Figure 1. Flowchart of the improved algorithm.

## 3. The Application of the HS-BP Model in Workers' Behavioral Risk Assessment

3.1. Data Collection and Sources

This paper uses a questionnaire to collect data and questionnaire background information.

Zhangzhou Nuclear Power is currently in the construction stage, with many construction units at the construction site, a wide distribution of construction risk points, and more than 20,000 staff in the region. The high social concern and strict safety requirements of the nuclear power plant under construction make it specific to both the high safety level requirements of a nuclear power plant and the complexity of the types of risks of a construction enterprise. Therefore, a safety supervision grid has been laid out for a number of key areas, requiring plenty of effort to work and patrol the monitoring.

The questionnaires were sent to managers, technicians, and frontline operators. A total of 1500 paper questionnaires were distributed, and 1481 were recovered, with a recovery rate of 98.73%. Discounting missing questions and invalid questionnaires that do not meet the requirements, the total number of valid questionnaires was 1397, with an effective questionnaire rate of 93.13%.

The questionnaire results and the summary results of the basic information of the respondents are shown in Figure 2.
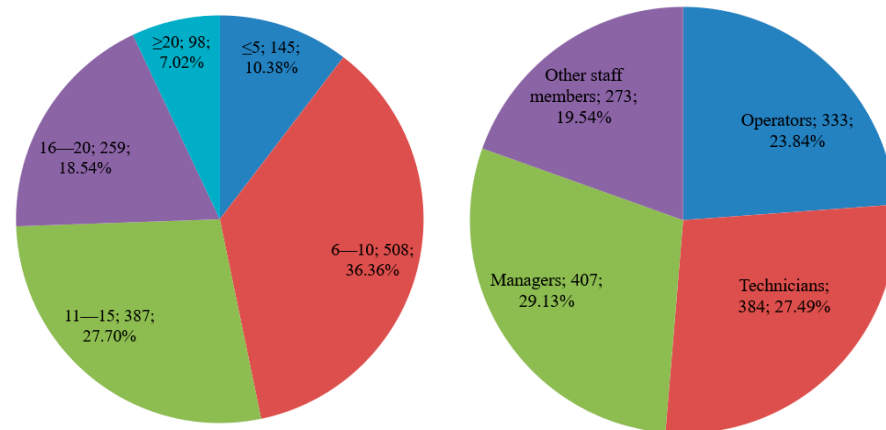
Figure 2. Distribution of respondents by length of service and position.

Analyzing the statistical characteristics of the basic information of the respondents, it can be seen that the group of respondents of this research, i.e., the sample, is reasonably structured, and the statistical information of the respondents in terms of their length of service, position, etc., is in line with the reality of the enterprise, which effectively ensures the objectivity of the survey results.

When dealing with some large dataset modeling problems, the original data cannot be adapted to the modeling needs due to a series of problems such as rich data sources, diverse data formats, and a lack of data completeness, so the dataset is generally subjected to preprocessing work before the formal model is built. Some of the data for the employee behavioral assessment in nuclear power plants under construction are entered manually on a daily basis, and there are irregularities or missing data records, so, in order to avoid the influence of noise on the final evaluation results, the data need to be preprocessed. In addition, in order to ensure the accuracy of the modeling, the collected variables need to be screened to select the indicators that have a greater impact on the final results.

Data pre-processing usually refers to checking the accuracy of the dataset and grasping its basic situation before formally establishing the model and transforming the raw data into a suitable form of data for modeling through a series of operations such as data cleansing, data integration and transformation, and a data statute in order to solve the problems of invalidity, high concentration, missing values, outliers, and inconsistencies of the data. The data in this paper come from the record of unsafe behaviors in the field, which is mainly obtained from the risk management platform, examination system, attendance management system, and monitoring and surveillance system. Part of the data is through questionnaires, and Table 1 gives some of the data for model training.

Primary variables: X_1 is age; X_2 is working age; X_3 is education level; X_4 is attendance; X_5 is psychological condition; X_6 is knowledge and skills; X_7 is safety awareness; X_8 is noise; X_9 is temperature; X_{10} is the concentration of welding fume; X_{11} is the concentration of construction dust; X_{12} is supervision and inspection; X_{13} is work arrangement; X_{14} is

safety education; X_{15} is emergency management; X_{16} is safety warnings; X_{17} for equipment reliability; and X_{18} for protective gear.

Unlike the risk assessment of mechanical and electrical equipment, the unsafe behaviors of operators cannot be directly measured but need to be measured with the help of their external manifestations (e.g., unauthorized operation during operation, the violation of labor discipline, etc.) Due to the externality of the above behaviors or conditions, it is easy to obtain the data, and, therefore, the concept of observing the risk level is proposed. The collected records of unsafe behaviors are classified according to the severity of the consequences of the unsafe behaviors: they are divided into four categories: general unsafe behaviors (I), more serious unsafe behaviors (II), serious unsafe behaviors (III), and accidental unsafe behaviors (IV). The results of the risk assessment of unsafe behaviors of employees of nuclear power plants under construction reflect the estimated value of the steady state point of the evaluated object during the work period, without taking into account the interval estimation of its periodic or transient fluctuations.

3.2. Indicator Screening

A total of 19 indicators were selected in the evaluation index fitting, some of which may have less influence on unsafe behaviors, and the fitting effect may not be satisfactory if 19 variables are used blindly to build the risk assessment model of employees' unsafe behaviors without screening. Therefore, in order to reduce the coupling effect between various factors, the important factors affecting unsafe behaviors were extracted through the random forest algorithm, and the data in Table 3 were used for analysis and calculation.

Table 3. Raw data for the evaluation of staff behavior.

X_1	39	31	31	31	31	36	35	36	40	30	28
X_2	5	4	3	1	4	2	3	3	2	1	2
X_3	1	1	2	3	1	2	2	1	1	3	2
X_4	28	29	25	27	22	28	25	28	30	28	28
X_5	78.1	81.4	73.8	73.8	95.7	65.4	90.1	81.6	84.5	94.3	96.0
X_6	86	83	86	85	83	81	82	83	82	75	76
X_7	86	82	93	82	84	86	86	89	85	81	89
X_8	81	80	75	89	86	87	80	81	83	84	83
X_9	24.58	22.47	20.53	23.19	25.19	25.52	23.05	22.13	23.26	22.78	23.5
X_{10}	0.01	0	0	0	0.02	0.03	0	0	0	0	0
X_{11}	0.92	0.02	0.04	0.06	1.59	2.89	0.11	0.02	0.02	0.18	0.02
X_{12}	86	89	82	84	81	83	84	85	82	87	88
X_{13}	88	89	86	84	88	86	85	88	89	91	86
X_{14}	84	86	71	91	86	83	79	85	88	76	72
X_{15}	95	94	95	96	95	95	94	96	97	92	93
X_{16}	81	82	84	81	81	83	84	85	88	83	82
X_{17}	80	81	83	82	81	85	88	83	82	81	82
X_{18}	82	84	82	80	86	82	81	80	84	83	79
X_{19}	82	84	82	80	86	82	81	80	84	83	81
X_{20}	I	II	II	III	II	II	III	II	II	II	III

3.3. Correlation Analysis of Variables

Before establishing the model, the correlation between the feature variables is also verified. If the feature variables have strong correlation, it will have a certain degree of impact on the generalization ability of the model. The correlation coefficient heatmap is

drawn for each characteristic variable studied in this paper to observe the correlation degree of each characteristic variable. Firstly, the correlation coefficient between each characteristic variable is calculated by using the Corr function, then, the correlation coefficient heatmap is drawn by using the heatmap function in the seaborn plotting package in Python, as shown in Figure 3, and the correlation coefficient is calculated by using the heatmap function in the seaborn plotting package in Python.
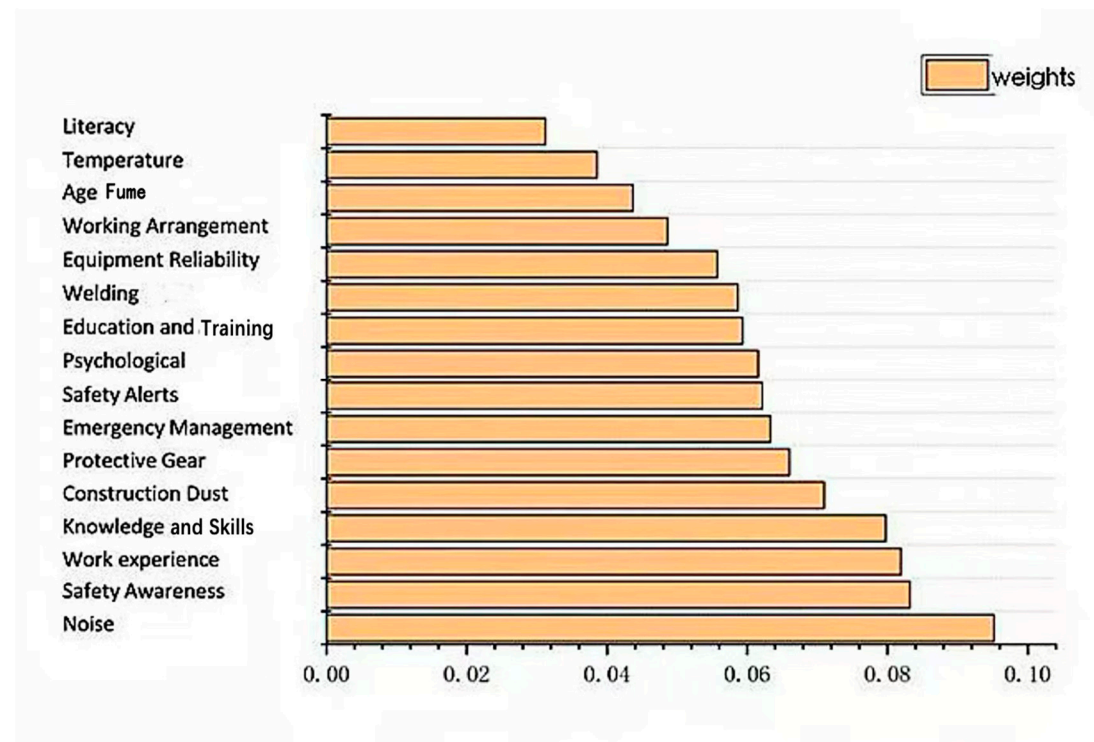
Figure 3. The extent to which indicators influence unsafe behavior.

According to the screening of feature variables of random forest, the final selection of age, working age, psychological factors, safety awareness, knowledge and skills, noise, temperature, work arrangement, education and training, supervision and inspection, safety warning, protective gear, equipment reliability, and the number of nodes determined in the input layer of the BP neural network is 13.

3.4. Network Model Setup

It was shown that a three-layer BP neural network can meet the fitting accuracy of any nonlinear function. Therefore, the model for this study was chosen to contain only one hidden layer. There exists an optimal number of nodes in the hidden layers, but how to determine it mainly depends on empirical formulas and experiments.

$$k = \sqrt{n + m} + a \tag{5}$$

where n is the number of nodes in the output layer; m is the number of nodes in the input layer; and a is an integer in the range 1–10.

The optimal number of nodes in the hidden layer is obtained by bringing $n = 4$ and $m = 12$ into the empirical Equation (5). The calculation results show that the number of nodes lies between the interval [5,14], and each integer of the various kinds is tested one by one with the BP neural network. The results of the training error are shown in Figure 4.

As can be seen in Figure 5, the error is minimized when the number of nodes in the hidden layer is 13. Therefore, the model in this study has 13 nodes in the input layer, 4 nodes in the output layer, and 10 nodes in the hidden layer.
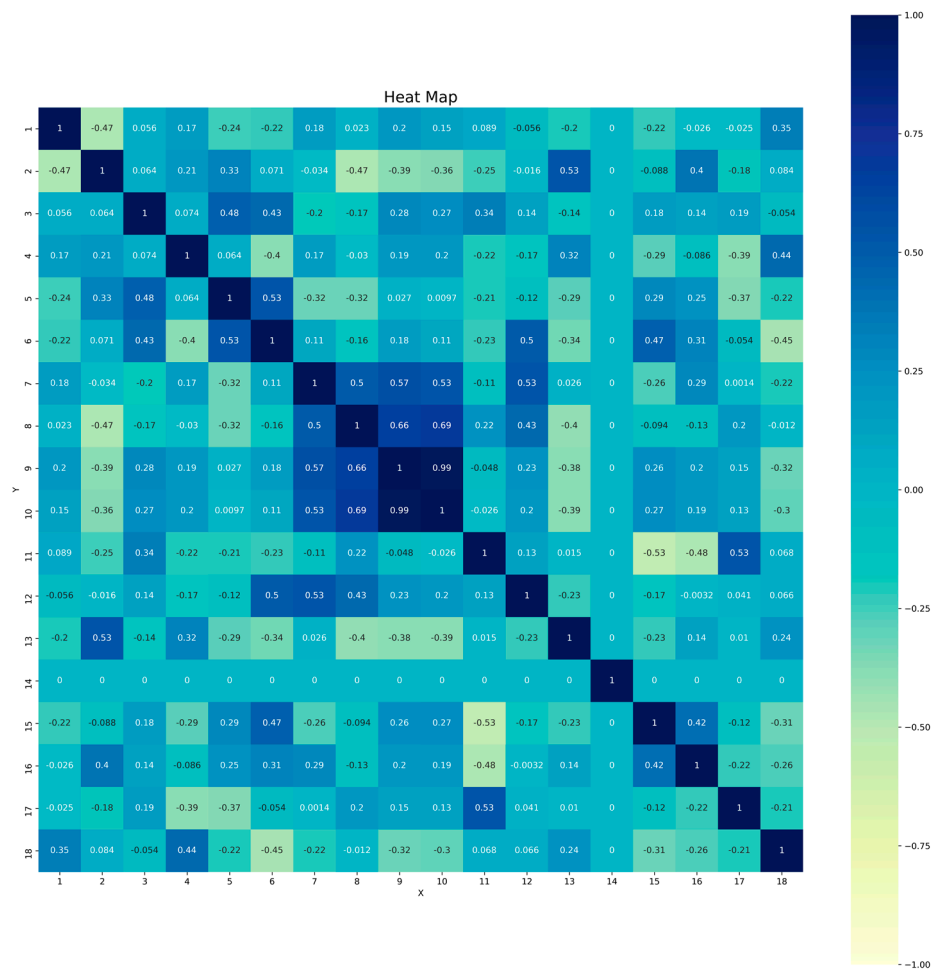
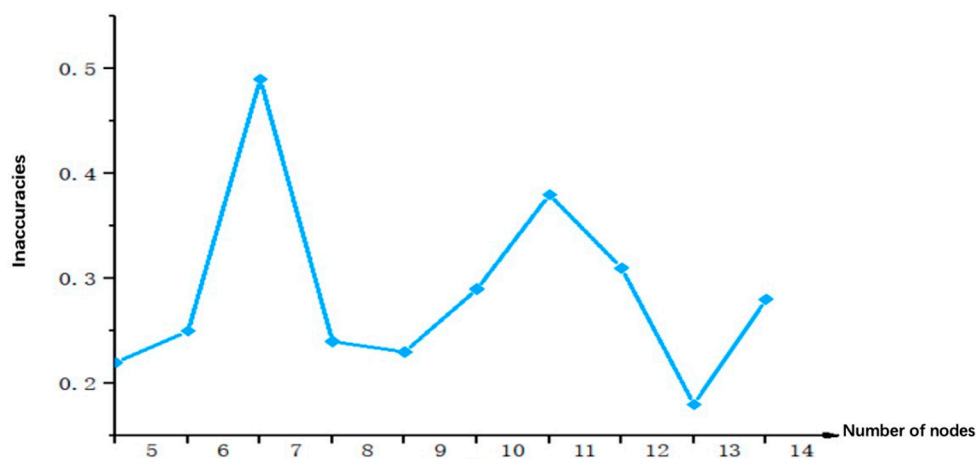**Figure 4.** Correlation coefficient heat map.

Figure 5. Training error of hidden nodes.

3.5. HS-BP Neural Network

The weight thresholds of BP neural networks are usually limited to a certain range. If the data are not normalized, the input range will be large. The corresponding range of the weight threshold adjustment will also become large, increasing the difficulty of adjusting the weight threshold. Therefore, in order to improve the convergence of the algorithm, it is necessary to normalize the inputs to [0, 1], and the transformation formula is Equation (6).

$$X_i^{'} = (X_i - X_{\min})/(X_{\max} - X_{\min}) \tag{6}$$

where X_i denotes the input data; X_{max} denotes the maximum value in the desired range; X_{min} denotes the minimum value in the desired range; and X_i' denotes the transformed input data.

The experiment was carried out using Python programming with a total of 313 datasets, of which 300 were used for training and 13 for testing.The HS-BP neural network model parameters are shown in Table 4.

Table 4. HS-BP neural network model parameters.

Memory Bank Size HMS	Memory Value Probability HMCR	Tuning Probability PAR	Tuning Step	Number of Nodes in the Hidden Layer of the BP Neural Network	Learning Rate	Number of Iterations
100	0	0.3	0.2	13	0.01	500

The two algorithms that model the BP model and HS-BP model use the same network structure with 70% of the 300 training samples as training and 30% as validation samples. After each round of learning, the network mean square error is obtained, and the maximum number of learning rounds is set to 500. The training error of both algorithms is shown in Figure 6.
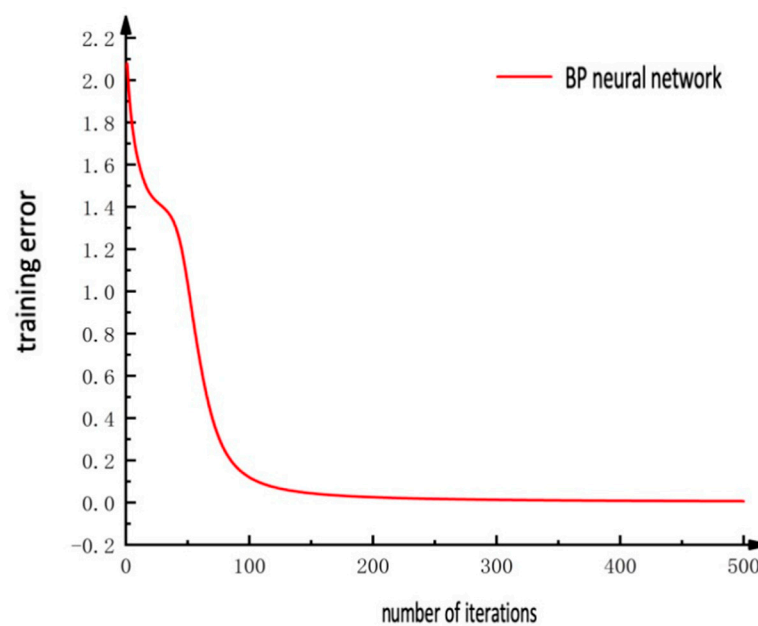
Figure 6. Training error plots for both algorithms.

As can be seen from the error curve, for the HS-BP model, after 203 rounds of data learning, the net error is reduced to 0.000805, which reaches the preset error accuracy of 0.001. Then, the learning stops. Despite the minor fluctuations in the error reduction process, the results are still in line with the expected state of the HS algorithm. The algorithm is very stable with no error stagnation. For the BP model, the network error decreases with significant fluctuations. Error stagnation occurs when the error is 0.2 and 0.03. After 400 rounds of learning, the average network error fluctuates around 0.017. Even when the maximum training number of 500 is reached, the network training error accuracy cannot reach the target value of 0.001, and the final error is 0.0172.

The remaining 13 data were compared by the advantages of the two algorithmic models, as shown in Figure 7. It is clear from this figure that the HS-BP algorithm model has the highest fit to the expected output.

Figure 7. Line graph of two algorithmic models.

The research methodology system was experimentally applied in Zhangzhou Nuclear Power, mainly realizing behavioral risk control and the multi-scenario hidden danger investigation for 20,000 workers in the under-construction nuclear power plant. The application platform is shown in Figure 8, which displays the risk and hidden danger data of the whole working area in real time.

Figure 8. Application platform diagram * （Some data sensitive to nuclear power are processed）.

4. Results and Discussion

In this study, the factors affecting the unsafe behaviors of employees in nuclear power plants under construction were identified through the random forest method, and

the input variables of the assessment model were determined on this basis. The HS-BP model was used to achieve the assessment of the risk of employees' unsafe behaviors, and, by comparing it with the BP model, the HS-BP model has high accuracy and convergence speed. The unsafe risk level of the workers in the nuclear power plant under construction is determined, and the personnel who need to be focused on supervision are identified, which provides a basis for the enterprise personnel management. The established assessment model takes into account the human self-factors, environmental factors, organizational factors, and mechanical equipment factors, thus reflecting to a certain extent their influence on the unsafe behaviors of the workers. The results show that the model has a good fitting effect.

In addition, in screening the importance of the factors influencing unsafe behaviors using random forests, we found that safety awareness (0.0832), working age (0.0819), and knowledge and skills (0.0797) have a greater influence on unsafe behaviors. The main reason for this is that workers in nuclear power plants under construction need more operational skills and experience to cope with the complex operating environment. Only by improving the safety awareness of workers and continuously strengthening education and training can the unsafe behavior of workers be effectively reduced and the occurrence of accidents in under-construction nuclear power plants be reduced. Among the factors influencing unsafe behaviors, noise has the greatest influence on unsafe behaviors (0.0952), and companies should reduce the exposure time of workers to noise and improve their awareness of noise prevention.

Accidental unsafe behaviors are affected by a combination of factors, and some factors (e.g., safety attitudes and safety culture) were not included in the evaluation system because they are not easily accessible or difficult to quantify when screening unsafe behavior influencing factors. In future studies, it is planned that appropriate databases will be established to store such data. A more detailed evaluation system will then be developed for a more accurate assessment.

5. Conclusions

Based on the BP neural network optimized by the random forest and harmony search algorithms, a risk assessment model for unsafe behaviors of employees in nuclear power plants under construction was established, and a systematic assessment of unsafe behaviors of employees was carried out to determine the level of unsafe behaviors. The main research conclusions are as follows:

(1) The feature variables affecting unsafe behaviors are screened by the random forest model, and, then, a BP neural network-based model is constructed to evaluate the screened feature variables, which can comprehensively and effectively analyze the unsafe behaviors of employees in nuclear power plants under construction.

(2) The BP neural network is optimized by using the harmonic search algorithm, and an assessment model of the unsafe behaviors of employees in nuclear power plants under construction is constructed, which is compared with the BP neural network model before optimization, and shows that the optimized model has higher accuracy.

(3) The employee behavior assessment based on the machine learning method outputs the risk level of employee behavior in nuclear power plants under construction more objectively and solves the deficiencies of one-sidedness and subjectivity of the traditional expert evaluation method.

Focus on key influencing factors, strengthen safety awareness education, and carry out differentiated knowledge and skill training for workers of different working ages; control environmental influencing factors, adopt noise reduction measures and strengthen workers' education on noise prevention; improve the data and assessment system, setup

a database for storing data on safety attitudes and other factors that are not easily quantifiable, and build a more detailed assessment system; and combine the management of assessment results with the management of risk levels to classify and manage the workers according to their risk levels, with high-risk specialized staff to focus on the supervision of high-risk workers, and regular reminders to strengthen safety awareness.

Future work around the risk assessment of unsafe behaviors of employees in nuclear power plants under construction can be carried out in many ways. In terms of model optimization and expansion, try to combine intelligent algorithms such as particle swarm and genetics with BP neural networks to compare performance while optimizing the structure of BP neural networks, exploring their advantages of combining them with deep learning models such as convolutional and recurrent neural networks and random forests, and integrating multi-source data such as employee physiology, operating behaviors, and historical accident cases. In terms of data, deep learning data interpolation and other techniques are studied to improve data quality, establish a long-term monitoring mechanism to update data regularly, and promote data sharing and comparison with other nuclear power plants and related industries. In terms of practical application, formulate and optimize interventions for the personnel of different risk levels, develop a real-time risk warning system, and extend the model to high-risk industries such as chemicals and mines. Theoretical research should deeply analyze the synergistic mechanism of various influencing factors and establish a standard system of risk assessment by combining research and application experience, so as to provide a normative reference for the industry.

This paper identifies personnel with higher behavioral risk levels through the assessment of employee behavior in under-construction nuclear power plants, which provides a basis for personnel management and has obvious practical significance and theoretical value. In addition, the influencing factors of unsafe behaviors were screened before the assessment was conducted, which can provide reference for other studies.

References

1. Reason, J. *Human Error*; Cambridge University Press: Cambridge, UK, 1990.
2. Dhillon, B.S. *Human Reliability: With Human Factors*; Elsevier: Amsterdam, The Netherlands, 2013.

3. Ming, X.; Yu, G.; Deng, S. HFE application in human system inter-face design of nuclear power plant. In Proceedings of the International Conference on Applied Human Factors and Ergonomics, Los Angeles, CA, USA, 17–21 July 2017; Springer: Los Angeles, CA, USA, 2017; pp. 361–370.

4. O'hara, J.; Higgins, J.; Flegers, S. Human factors engineering program review model, NUREG-0711; U.S. Nuclear Regulatory Commission: Washington, DC, USA, 2012.

5. Zhang, L.; Wei, H. Experimental study on the data collecting errors with the operator's monitoring of the digitalized human-machine interface. *J. Saf. Environ.* **2016**, *16*, 191–195.

6. Fink, R.; Hill, D.; O'hara, J. *Human Factors Guidance for Control Room and Digital Human-System Interface Design and Modification, Guidelines for Planning, Specification, Design, Licensing, Implementation, Training, Operation and Maintenance*; Electric Power Research Institute: Palo Alto, CA, USA, 2004.

7. Higgins, J.C.; O'hara, J.M.; Lewis, P.M.; Persensky, J.J.; Bongarra, J.P.; Cooper, S.E.; Parry, G.W. *Guidance for the Review of Changes to Human Actions, NUREG-1764*; Nuclear Regulatory Commission: Washington, DC, USA, 2007.

8. Higgins, J.C.; O'hara, J.M.; Lewis, P.M.; Persensky, J.; Bongarra, J. Development of a risk screening method for credited operator actions. In Proceedings of the 7th Conference in Human Factors and Power Plants, Scottsdale, Arizona, 15–19 September 2002; Institute of Electrical and Electronic Engineers: Washington, DC, USA, 2002.

9. Westinghouse Electric Company. *AP1000 Identification of Critical Human Actions and Risk Important Tasks, WCAP-16555*; Westinghouse Electric Company: Pittsburgh, PA, USA, 2006.

10. Zheng, H. *Probabilistic Risk Assessment*; National Defense Industry Press: Beijing, China, 2011.

11. Zhao, J.; Tong, J.; He, X. Risk-informed SSC categorization method in nuclear powerplant. *J. Tsinghua Univ. Sci. Technol.* **2007**, *47*, 449–452.

12. Choi, B.; Ahn, S.; Lee, S. Construction Workers' Group Norms and Personal Standards Regarding Safety Behavior: Social Identity Theory Perspective. *J. Manag. Eng.* **2017**, *33*, 04017001.

13. Deng, Y.; Guo, H.; Meng, M.; Zhang, Y.; Pei, S. Exploring the Effects of Safety Climate on Worker's Safety Behavior in Subway Operation. *Sustainability* **2020**, *12*, 8310.

14. Fernández-Muñiz, B.; Montes-Peón, J.M.; Vázquez-Ordás, C.J. The role of safety leadership and working conditions in safety performance in process industries. *J. Loss Prev. Process Ind*. **2017**, *50*, 403–415.

15. Mokarami, H.; Alizadeh, S.S.; Pordanjani, T.R.; Varmazyar, S. The relationship between organizational safety culture and unsafe behaviors, and accidents among public transport bus drivers using structural equation modeling. *Transp. Res. Part F Traffic Psychol. Behav.* **2019**, *65*, 46–55.

16. Zhang, J.; Fu, J.; Hao, H.; Fu, G.; Nie, F.; Zhang, W. Root causes of coal mine accidents: Characteristics of safety culture deficiencies based on accident statistics. *Process Saf. Environ. Prot*. **2020**, *136*, 78–91.

17. Ye, G.; Duan, S.; Wang, H. Research on the Causes of Construction Workers' Unsafe Behavior. *China Saf. Prod. Sci. Technol.* **2015**, *11*, 170–177.

18. Cheng, J.; Qi, S.; Zhang, Y. Mechanism and empirical research on the influence of organisational climate on construction workers' unsafe behaviors. *China Saf. Prod. Sci. Technol.* **2017**, *13*, 11–16.

19. Zhang, H.; Qi, C.; Ma, M. Improved Employee Safety Behavior Risk Assessment of the Train Operation Department Based on Grids. *Processes* **2022**, *10*, 1162.

20. Hong, Y.; Cho, J. Enhancing Individual Worker Risk Awareness: A Location-Based Safety Check System for Real-Time Hazard Warnings in Work-Zones. *Buildings* **2024**, *14*, 90. https://doi.org/10.3390/buildings14010090.

21. Yu, K.; Zhang, S.; Liu, P.; Zhou, L.; Chen, J. Research on Simulation and Decision-Making of Coal Mine Workers' Behavior Risk. *Mathematics* **2022**, *10*, 3297. https://doi.org/10.3390/math10183297.

22. Zhang, S.; Cao, Q.; Jia, H. Research on the application of EAP system to enhance the safety psychology level of coal mine workers. *China Coal* **2016**, *42*, 10–15. https://doi.org/10.19880/j.cnki.ccm.2016.01.003.

23. Cao, Y. Research on the propagation and simulation of unsafe behaviors of civil aviation flight attendants. *Civ. Aviat. Univ. China* **2022**. https://doi.org/10.27627/d.cnki.gzmhy.2022.000269.

24. Maslow, A.H. *The Farther Reaches of Human Nature*; Bassett, M., Ed.; 1973.

25. Bandura, A. Chapter 3-Modeling Theory: Some Traditions, Trends, united states Public health Services. In *Recent Trends in Social Learning Theory*; Parke, R.D., Ed.; Academic Press: Cambridge, MA, USA, 1972; pp. 35–61. ISBN9780 125450508.

26. Vroom, V.H.; Jago, A.G. *The New Leadership: Managing Participation Inorganizations*; Prentice-Hall, Inc.: Upper Saddle River, NJ, USA, 1988.

27. Fishbein, M.; Ajzen, I. *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*; Addison-Wesley Publishing Company: Reading, MA, USA, 1975; p. 53.

28. Greenwood, M.; Woods, H.M. *The Incidence of Industrial Accidents upon Individuals: With Special Reference to Multiple Accidents*; H M Stationery Office: Norwich, UK, 1919.

29. Heinrich, H.W. Industrial Accident Prevention. A Scientific Approach. *Ind. Accid. Prev. A Sci. Approach* **1959**, *4*, 609.

30. Surry, J. *Industrial Accident Research: A Human Engineering Appraisal*; University of Toronto: Toronto, ON, Canada, 1969.

31. Reason, J. Human error. *West J. Med.* **1990**, *12*, 393–396.

32. Cao, Q.; Li, S.; Song, X. Research on the 'Knowing-Capable-Acting' unsafe behavior pattern of coal mine employees. *China J. Saf. Sci.* **2007**, *117*, 19–25.

33. Zhou, P. *MATLAB Neural Network Design and Applications*; Tsinghua University Press: Beijing, China, 2013. Zhou, P. *Design and Application of MATLAB Neural Network*; Tsinghua University Press: Beijing, China, 2013.

34. Xia, Y.; Wang, L. Back propagation neural network based bidding risk evaluation method for marine engineering projects. *J. Shanghai Jiao Tong Univ.* **2023**, *57*, 46–53. Xia, Y.; Wang, L. Evaluation method of bidding risk of offshore engineering project based on back propagation neural network. *J. Shanghai Jiao Tong Univ.* **2023**, *57*, 46–53.

35. Dong, J.W.; Hu, J.M.; Luo, Z. Quality prediction of aluminium alloy resistance spot welding based on correlation analysis and SSA-BP neural network. *J. Weld.* **2024**, *45*, 13–18.

36. Gao, Y.; Chen, D. Development of cable fault location equipment based on low-voltage pulse method. *Sci. Innov.* **2024**, 132–134.

37. Zhou, Z.; Lin, D. Early warning of fault location of substation equipment based on digital twin technology. *Comput. Program. Ski. Maint.* **2023**, 173–176.

38. Cao, H.-O.; Wu, D.; Xue, F.; Wang, Y.; Sun, H.; Yang, J. Fault location method of secondary equipment in intelligent substation based on improved BA-PNN. *Intell. Power* **2024**, *52*, 32–39.