IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 28 July 2024, accepted 10 August 2024, date of publication 15 August 2024, date of current version 28 August 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3444485*

RESEARCH ARTICLE

# Research on Intelligent Diagnosis and Decision-Making Method for Oilfield Water Injection System Faults

**RUIJIE ZHANG[1,2], WENTING YANG[1], JIE LI[1], SHENGLIANG GAO[1], YAN WANG[1,2], AND SHENG GAO[1,2]**

[1]School of Mechanical Science and Engineering, Northeast Petroleum University, Daqing 163318, China
[2]Laboratory of System Simulation and Control, Northeast Petroleum University, Daqing 163318, China

Corresponding author: Ruijie Zhang (261974020133@nepu.edu.cn)

This work was supported by U.S.-China Clean Energy Research Centre Joint Work Plan for Research Projects on Water Energy Technologies: Non Traditional Water Resources Treatment and Management under Grant 2018YFE019600.

**ABSTRACT** Water injection is a commonly used development method in oilfields. Water injection systems are large and complex, with horizontally and vertically interconnected pipeline networks often buried underground. Faults in pipeline networks cannot be detected and handled in time, thereby posing significant safety hazards to production. This study focuses on fault diagnosis and decision-making within a water injection system. We established a fault tree for an oilfield water injection system and proposed an optimized BP neural network with a Self-Adaptive Differential Evolution Algorithm for the first time. This method constructs a two-layer fault diagnosis model for a water injection system. The model diagnosed fault positions and types based on parameters such as the fault point flow, pressure, and pipeline flow. Compared with the traditional BP algorithm, this algorithm has better diagnostic accuracy and faster convergence speed. Simultaneously, the decision tree CART method was employed to classify decision types based on multiple parameter indicators of the fault points and generated decisions. We designed and implemented a fault diagnosis and decision platform for an oilfield water injection system. Finally we built an experimental pipeline network model with EPANET to simulate the system fault conditions. The diagnostic performance of the proposed algorithm was tested. The results showed that the proposed method achieved a 99% accuracy rate in diagnosing faults in a water injection system. This method significantly improves the scientific management of water injection systems, holding great potential for broad application and value in achieving smart oilfields.

**INDEX TERMS** Oilfield water injection system, SDE-BP algorithm, fault diagnosis, CART, decision-making.

## I. INTRODUCTION

Water-flooding displacement is one of the most commonly used development methods in oil fields, which requires a water injection system to inject water into the oilfield formation under high pressure and flow rate. The water injection system is a large, enclosed, and complex hydrodynamic system consisting of water injection stations, water distribution rooms, water injection wells, and water injection pipelines (see Fig. 1) that are buried underground. The system runs continuously for 24 h a day, consuming a large amount of energy. Because pipelines are buried underground, they are often affected by water quality corrosion, water hammer, and the geological environment, resulting in faults such as pipeline jets and bursting. Currently, regular inspection and maintenance of pipeline networks often consumes considerable manpower and material resources. In the face of sudden events the emergency warnings and effective response

The associate editor coordinating the review of this manuscript and approving it for publication was Lei Shu.

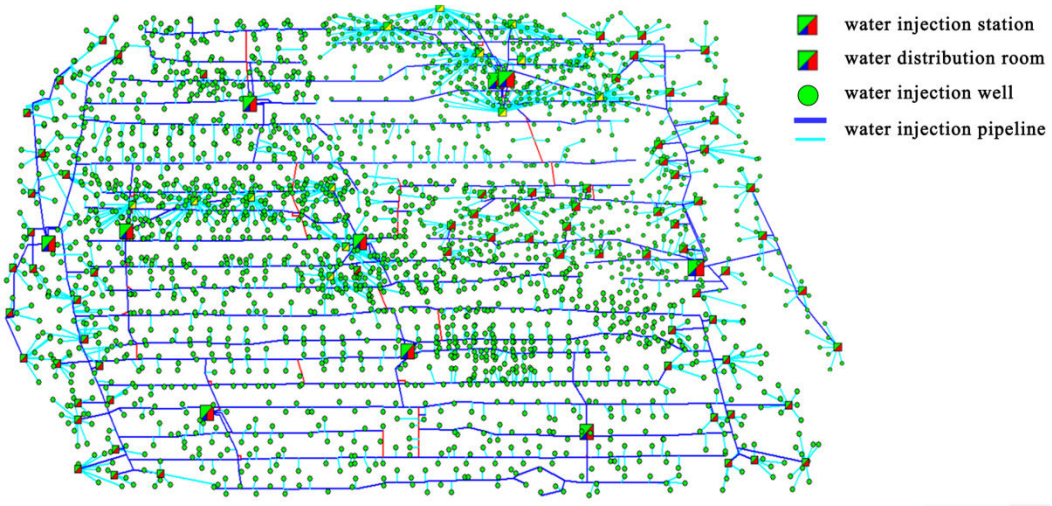


**FIGURE 1.** The structure diagram of a water injection system in Daqing oilfield.

capabilities are lacking [1]. Moreover, the running status of the water injection pipeline network changes dynamically with oilfield production, and the manual management mode can no longer meet the demand for dynamic production. Therefore, it is necessary to fully utilize the large amount of data generated during the system operation to analyze the water injection working status and perform intelligent diagnosis for faults in water injection system and decision-making, which will enable early warning and timely maintenance, avoid large-scale incidents [2], and therefore enhance the intelligent management level in oilfield water injection [3].

Little research has been conducted on fault diagnosis and decision-making for oilfield water injection systems, both domestically and internationally. The fault diagnosis techniques employed in urban water supply networks can be [4], [5]. In recent years, various intelligent algorithms have been incorporated into research on fault diagnosis in water supply and heating networks. Song Lanhua proposed an automated method for diagnosing faults in rural water supply systems. Based on fault classification, a Self-Organizing Map (SOM) neural network was introduced to generate a two-dimensional image result for fault diagnosis [6]. Qi Chengwei et al. introduced a water supply network burst diagnosis model based on HHT and SVM. The processed hydraulic data were utilized as the input for particle swarm optimization and support vector machine models [7], [8]. Weisong et al. proposed an improved BP neural network pipeline fault diagnosis method that utilize the GAAA algorithm. The leakage points and leakage amounts of the pipeline network can be determined By optimizing the initial weights and thresholds of the BPNN [9]. Ming proposed a leak diagnosis method for a central heating network based on deep learning. Using the TensorFlow deep learning framework, convolutional neural network (CNN) models were established to predict the leakage degree and location in the heating network [10]. In terms of auxiliary decision-making for water injection systems, Liu Shumeng et al. constructed an energy consumption- balancing model for a water injection system to analyze the spatial distribution, production data, and water injection energy consumption of the pipeline network. This model provides auxiliary decision support for the operation of a water injection system [11]. Yan Juan et al., starting with the digitalization of the water injection system, proposed the construction of a water injection station control system centered on equipment control and real-time monitoring. This system aims to achieve dynamic monitoring and analysis of a water injection system in order to enhance the automation level and scientific decision-making capabilities of oilfield water injection systems [12].

Aiming at fault diagnosis and decision-making challenges in oilfield water injection networks, with the goal of establishing a smart oilfield, a Differential Evolution algorithm is utilized to optimize neural networks, creating an adaptive differential neural network (SDE-BP) hybrid diagnostic model. The model performs the primary diagnosis of fault locations and secondary diagnosis of fault types. Concurrently, a decision tree for Classification and Regression Trees (CART) was constructed for decision partitioning. The EPANET software was chosen to build an experimental model for the water injection pipeline network. We developed a platform using Python, that can simulate system fault conditions and conduct experimental research on the second-level diagnosis of system faults.

## II. MATERIALS AND METHODS

### A. DESIGN OF FAULT TREE AND DIAGNOSTIC CRITERIA FOR WATER INJECTION SYSTEM

#### 1) ESTABLISHING FAULT TREE FOR WATER INJECTION SYSTEM

There are many faults in the oilfield water injection systems [13], [14], [15]. Faults in a water injection system refer


**FIGURE 2.** Fault tree of water injection system.

to hardware failures that occur in various parts of the pumping station, pipeline network, and injection wells. Environmental disturbances to the pipeline network mainly originate from the geological environment in which the pipeline network and injection wells are embedded. However, the geological environment does not directly cause abnormalities in the flow rate or pressure of the pipeline network.

Therefore, any abnormalities in the pressure and flow rate of the water injection system can be attributed to faults within the system itself. Through literature retrieval, thorough research, and consultation with authoritative experts in the oilfield industry, it was found that these faults are primarily distributed in three areas: water injection stations, water injection wells, and water injection pipelines. The specific classification of the fault types is shown in Fig. 2. This study only discussed fault diagnosis and decision-making technology related to fluid flow.

### 2) DESIGN OF FAULT DIAGNOSIS CRITERIA

In a water injection system, any fault related to fluid flow is reflected in the abnormal pressure and flow, which affects the surrounding water injection wells and pipeline segments. Therefore, the pressure and flow are crucial parameters for fault diagnosis. Typically, faults such as overload, underload, and stoppage of the pump unit in a station lead to abnormal pressure throughout the network. Similarly, faults such as over-injection, under-injection, and nozzle damage in water injection wells, as well as pipeline bursting, jetting, scaling, and blockage in the pipeline network, result in abnormal flow.

For the above faults, oil field sites typically provide general fault identification based on manual experience. In this paper, through investigation and research, a two-level fault diagnosis method based the SDE-BP model is proposed. First-level fault diagnosis determines the fault location based on pressure and flow anomalies. Pressure anomalies indicate the fault category for water injection stations, whereas flow anomalies indicate the fault categories for wells and pipelines. The second-level diagnosis determines the fault type based on pressure and flow anomaly thresholds. The preliminary diagnostic criteria for stations, wells, and pipelines are as follows [16].

- The high pressure at the water injection station indicates an overload of the station pump unit and faults in the pump outlet valve opening.

- The low pressure at the water injection station indicates an underload or stoppage of the pump unit.
- A high flow rate at a water injection well or a failure of the packer indicates over-injection at the water injection well.
- A low flow rate at the water injection well, clogged nozzles, and faults in the water injection valve opening indicate under-injection at the water injection well.
- No flow or low flow in the pipeline indicates blockage of the pipeline.
- The flow on the pipeline surface indicates three types of leakage faults: pipeline burst, jetting, and corrosion.

### B. INTRODUCTION OF THE SELF-ADAPTIVE DIFFERENTIAL EVOLUTION BACK PROPAGATION (SDE-BP) ALGORITHM

#### 1) FUSION STRATEGY OF DIFFERENTIAL EVOLUTION ALGORITHM AND NEURAL NETWORK

The Back Propagation (BP) neural network has been extensively utilized in numerous fault diagnosis applications [17], however its application in diagnosing faults in oilfield water injection systems remains rare. The primary impediment is the sluggish convergence rate of the traditional BP neural networks, particularly when dealing with extensive training sets. Furthermore, these networks are easy to overfit, resulting in a local optimum for model settling. However, the optimizing BP neural networks can effectively address these challenges [18], [19]. The proposed SDE-BP algorithm offers two significant advantages.

- In the adaptive differential evolution algorithm, by introducing adaptive mutation and crossover factors, the diversity of the population can be ensured in the early stage and good individuals can be retained in the later stage.
- Owing to the challenges associated with optimizing the input/output and hidden layer structures, a self-adaptive differential evolution algorithm was employed to optimize the weights and threshold parameters of the neural network. This approach compensates for inherent network deficiencies, enhances the training efficiency, and prevents the model from converging to local optima.

The SDE-BP model initializes all weights and thresholds as the initial population for the self-adaptive differential evolution algorithm. The sum of the weights and thresholds was considered as the length of the population. Through multiple adaptive mutation, crossover, and selection operations, optimized weights and thresholds were obtained, transforming them from an initial random state to a directionally optimized state. This accelerated the convergence of the BP network model. Based on this, a three-layer BP neural network was established. A diagnostic model of the SDE-BP network can be obtained by training [20].

**TABLE 1. Summary of mathematical notations.**

| Symbol | Description |
|---|---|
| $N_{\text{hid}}$ | The number of neurons in the hidden layer |
| $N_{\text{in}}$ | The number of neurons in the input layer |
| $N_{\text{out}}$ | The number of neurons in the output layer |
| $\alpha$ | The constant, $0 \leq \alpha \leq 10$ |
| rand | The random number between [0,1] |
| $w_j^{\max}$ | The maximum values of the initial weights. |
| $w_j^{\min}$ | The minimum values of the initial weights. |
| $G$ | The current evolutionary generation count |
| $\mathbf{W}_{best}^{G}$ | Excellent weights in the $G$-th generation |
| $r2$, $r3$ | Individuals in the population that are different from $i$ |
| $F_{\min}$ | The minimum mutation factor |
| $F_{\max}$ | The maximum mutation factor |
| $G_{\max}$ | The maximum iteration count |
| $F$ | values between 0.3 and 0.7 |
| $u_i^{G+1}$ | Tthe $i$-th individual in the $G+1$-th generation after crossover |
| $v_i^{G+1}$ | The $i$-th individual after mutation |
| $w_i^{G}$ | The original individual |
| $C_{\min}$ | The minimum mutation factor |
| $C_{\max}$ | The maximum mutation factor |
| $C$ | values between 0.1 and 0.9 |
| $f$ | The mean squared error evaluation function |
| $\mathbf{W}_i^{G+1}$ | The optimized weights for the new generation |
| $\mathbf{W}_{nm}^{(1)}$ | The $m \times n$ weight matrix from the input layer to the hidden layer |
| $\mathbf{W}_{mo}^{(2)}$ | The $o \times m$ weight matrix from the hidden layer to the output layer |
| $\text{sum}_k$ | The sum of weights for the $k$-th node in the hidden layer |
| $w_{ki}^{(1)}$ | The $i$-th row and $k$-th column element from the input layer to the hidden layer |
| $x_i$ | The $i$-th signal |
| $\theta_k^{(1)}$ | The $k$-th threshold from input layer to hidden layer |
| $z_k$ | The output value calculated through the activation function for the $k$-th node in the hidden layer |
| $\text{sum}_p$ | The weighted sum for the $p$-th node in the output layer |
| $w_{pi}^{(2)}$ | The elements in the $i$-th row and $p$-th column of weights from the hidden layer to the output layer |
| $z_i$ | The output signal of the $i$-th hidden layer node also serve as the input for the output layer |
| $\theta_p^{(2)}$ | The $p$-th threshold from the hidden layer to the output layer |
| $y_p$ | The output value calculated through the activation function for the $p$-th node in the output layer |
| $\partial$ | The learning rate |
| $\mathbf{E}_p$ | The error value |
| $t_p$ | The target value |
| $y_p$ | The output value from the output layer |
| $\mathbf{E}_p^{(2)}$ | The backward error from the output layer to the hidden layer |
| $\mathbf{E}_k^{(1)}$ | The backward error from the hidden layer to the input layer |
| $\mathbf{W}_{kp}^{T}$ | The transpose of the weight matrix $o \times m$ from the hidden layer to the output layer |
| $\Delta\mathbf{W}_{ik}$ | The weight correction value allocated to weights during the backward propagation of errors from the output layer to the hidden layer |
| $p_k$ | The proportion of the $k$-th sample in the sample set (D) |

### 2) SDE-BP FAULT DIAGNOSIS MODEL

The symbols and their descriptions in this study is shown in Table 1.

In the model, the pressure and flow of the pipeline network nodes were used as special diagnostic inputs, and the fault location and fault type were used as outputs. The number of hidden layer nodes is determined based on the number of network inputs and outputs. The number of hidden layer units was determined using the golden section method, with the formula given as: $N_{\text{hid}} = \sqrt{N_{\text{in}} + N_{\text{out}}} + \alpha$.

The BP neural network randomly generates initial weights and thresholds $w_1, w_2, w_3, \cdots, w_n, \theta_1, \theta_2$, The individual is $\mathbf{W}_i^G = (w_1, w_2, w_3, \cdots, w_n, \theta_1, \theta_2)$, $i = 1, 2, 3, \cdots$, NP, Each individual represents a set of solutions to the problem. This is used as the original individual for the SDE and is then optimized [21].

The optimization of the BP neural network weights and thresholds using the SDE algorithm is detailed as follows:

#### *a: POPULATION INITIALIZATION*

$$w_j = w_j^{\min} + \text{rand}(w_j^{\max} - w_j^{\min})\, j = 1,2,\ldots,n \tag{1}$$

#### *b: ADAPTIVE MUTATION OPERATION*

New individuals were generated between individual weights through mutations. The mutation operation uses a multi-differential improvement form to evolve the solution vector in a better direction, as shown in (2). The adaptive mutation operation introduces an adaptive mutation factor that decreases as the iteration count increases, as shown in (3).

$$\mathbf{V}_i^{G+1} = \mathbf{W}_{best}^G + F(\mathbf{W}_{r2}^G - \mathbf{W}_{r3}^G) \tag{2}$$

$$F = F_{\min} + (F_{\max} - F_{\min}) e^{1 - \frac{G_{\max}}{G_{\max} - G + 1}} \tag{3}$$

#### *c: ADAPTIVE CROSSOVER OPERATION*

The adaptive crossover operation is performed with the intermediate individual $\mathbf{V}_i^{G+1} = (v_{i1}^{G+1}, v_{i2}^{G+1}, v_{i3}^{G+1}, \cdots, v_{in}^{G+1})$ $i = 1, 2, 3, \cdots$ NP, obtained from the adaptive mutation operation and the original individual

$$\mathbf{X}_i^G = (w_{i1}, w_{i1}, w_{i3}, \cdots, w_{in}, b_{i1}, b_{i2}).$$

A candidate individual $\mathbf{U}_i^{G+1} = (u_{i1}^{G+1}, u_{i2}^{G+1}, u_{i3}^{G+1}, \cdots, u_{in}^{G+1})$ is obtained.

$$u_{im}^{G+1} = \begin{cases} v_{im}^{G+1} & \text{rand}(j) \le C \\ w_{im}^G & others \end{cases} \tag{4}$$

$$C = C_{\max} - \frac{G(C_{\max} - C_{\min})}{G_{\max}} \tag{5}$$

#### *d: SELECTION OPERATION*

Individuals $\mathbf{U}_i^{G+1}$ were selected for fitness evaluation. The evaluation function denoted by $f$ is the mean squared error of the output results, as shown in (6). Based on (7), it is decided whether to replace the current individual with the candidate individual in the next generation, thereby completing one cycle.

$$f = \frac{1}{n} \sum_{i=0}^{n} (t_i - \text{y}_i)^2 \tag{6}$$

$$\mathbf{W}_i^{G+1} \begin{cases} \mathbf{U}_i^{G+1} & f(\mathbf{U}_i^G) < f(\mathbf{W}_i^G) \\ \mathbf{W}_i^G \end{cases} \tag{7}$$

After multiple cycles of the Self-Differential Evolution (SDE) algorithm, the weights and thresholds can reach an optimal state, which is then combined with the pressure and flow of the pipeline network nodes to serve as the input feature values for the neural network.

#### *e: INITIALIZING THE NEURAL NETWORK*

A two-level SDE-BP network model was established to diagnose the fault locations and types in a water injection system. The model has a three-layer structure at both levels. In the first-level diagnostic model, the inputs include pressures $\mathbf{P} = (p_1, p_2, p_3, \cdots, p_n)$, flow rates $\mathbf{NQ} = (\text{nq}_1, \text{nq}_2, \text{nq}_3, \cdots, \text{nq}_n)$ of all nodes in water injection stations and wells, and segment pipeline flow rates $\mathbf{LQ} = (\text{lq}_1, \text{lq}_2, \text{lq}_3, \cdots, \text{lq}_m)$. In addition, the diagnostic classifica-tion labels $\mathbf{T} = (t_1, t_2, t_3, \cdots, t_k)$ were considered. The outputs represent fault locations. We established a second-level diagnostic model for each identified fault location $\mathbf{Y} = (y_1, y_2, y_3, \cdots, y_k)$, where the inputs were the pressure $\mathbf{P}$, flow rate $\mathbf{NQ}$ and pipeline flow rate $\mathbf{LQ}$, and the output was the fault type.

#### *f: NEURAL NETWORK FORWARD PROPAGATION TRAINING*

As each level and type of fault corresponds to a complete neural network model [22], the algorithm is explained uniformly. The input layer $(x_1, x_2, x_3, \cdots, x_n)$ consists of $n$ input nodes, the hidden layer has $m$ nodes, and the output layer has $o$ nodes. The thresholds are denoted as $\boldsymbol{\theta}^{(1)}$ $(\theta_1^{(1)}, \theta_2^{(1)}, \theta_3^{(1)}, \cdots, \theta_m^{(1)})$ and $\boldsymbol{\theta}^{(2)}$ $(\theta_1^{(2)}, \theta_2^{(2)}, \theta_3^{(2)}, \cdots, \theta_o^{(1)})$, the weights are represented as follows:

$$\mathbf{W}_{nm}^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{21}^{(1)} & \cdots & w_{m1}^{(1)} \\ w_{12}^{(1)} & w_{22}^{(1)} & \cdots & w_{m2}^{(1)} \\ \vdots & \vdots & \vdots & \vdots \\ w_{1n}^{(1)} & w_{2n}^{(1)} & \cdots & w_{mn}^{(1)} \end{bmatrix}$$

$$\mathbf{W}_{mo}^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{21}^{(2)} & \cdots & w_{o1}^{(2)} \\ w_{12}^{(2)} & w_{22}^{(2)} & \cdots & w_{o2}^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ w_{1m}^{(2)} & w_{2m}^{(2)} & \cdots & w_{om}^{(2)} \end{bmatrix}$$

The weighted sum of inputs is given by (8).

$$\text{sum}_k = \sum_{i=1}^{n} x_i^{(1)} w_{ki}^{(1)} + \theta_k^{(1)} \quad (k = 1, 2, 3, \cdots, m) \tag{8}$$

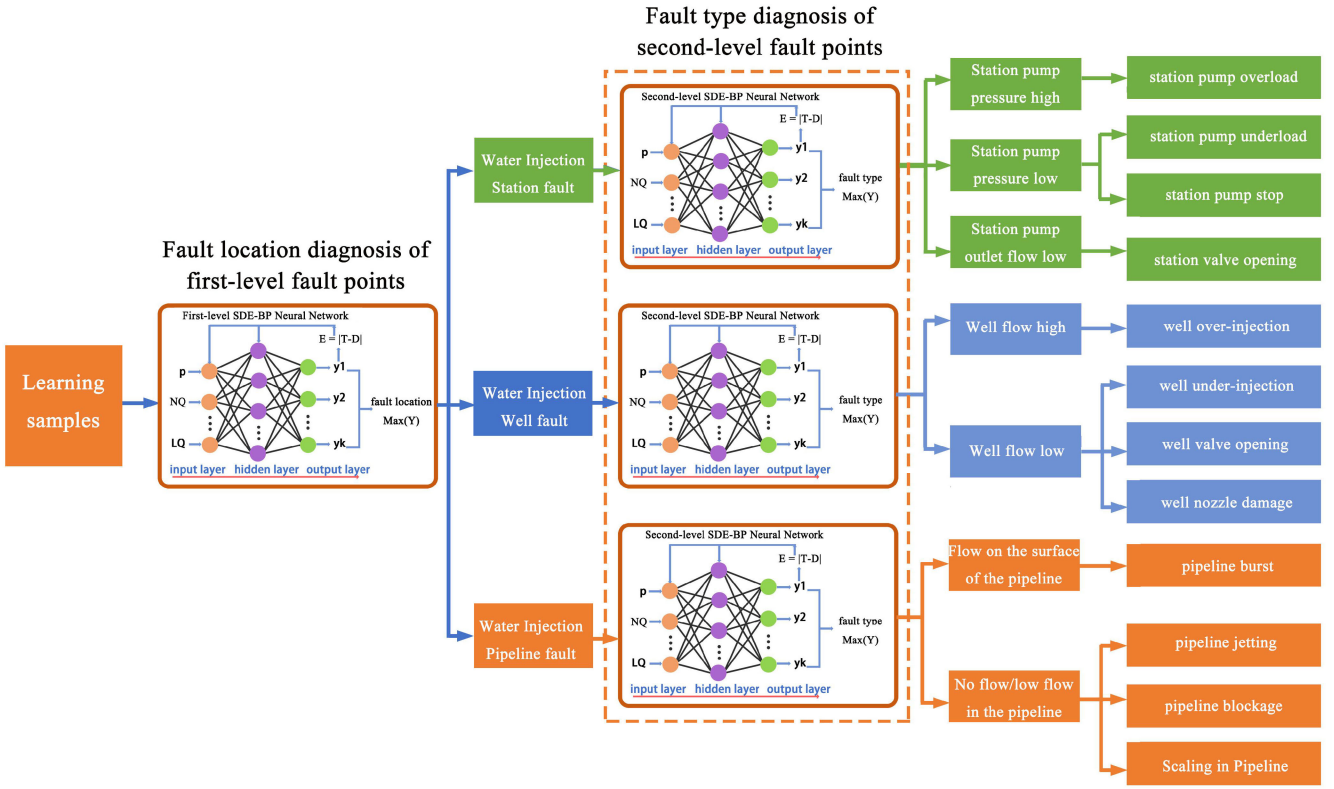


**FIGURE 3.** Overall architecture of the diagnostic model.

The results were introduced into the node activation function to calculate the output of the node. Here, the Sigmoid function is used, as shown in (9).

$$z_k = \text{Sigmoid}(\text{sum}_k) \quad (k = 1, 2, 3, \cdots, m) \tag{9}$$

The output result serves as the input for the next layer to calculate the weighted sum of the nodes. The final output based on the activation function is obtained, as shown in (10) and (11).

$$\text{sum}_p = \sum_{i=1}^{m} z_i w_{pi}^{(2)} + \theta_p^{(2)} \quad (p = 1, 2, 3, \cdots, o) \tag{10}$$

$$y_p = \text{Sigmoid}(\text{sum}_p) \quad (p = 1, 2, 3, \cdots, o) \tag{11}$$

### g: ERROR BACKWARD PROPAGATION TRAINING

As shown in (12), the error $\mathbf{E}$ is distributed across two layers, and the final layer does not perform error distribution

$$\mathbf{E}_p^{(2)} = \left| t_p - y_p \right| \quad (p = 1, 2, 3, \cdots, o)$$

The error distribution in this layer adopts a cross-entropy function, which drives the training to reduce the error more quickly, resulting in a faster learning process [23]. The error propagation and weight correction can be represented by matrices as shown in (13), (14) and (15), The learning rate $\partial$ can avoid generalization affecting the training results and also reduces the impact of the last data on the result.

$$\mathbf{E}_p = \left| t_p - y_p \right| \quad (p = 1, 2, 3, \cdots, o) \tag{12}$$

$$\begin{cases} \mathbf{E}_p^{(2)} = \left| t_p - y_p \right| & (p = 1, 2, 3, \cdots, o) \\ \mathbf{E}_k^{(1)} = \mathbf{W}_{kp}^T \mathbf{E}_p^{(2)} & (k = 1, 2, 3, \cdots, m) \end{cases} \tag{13}$$

$$\begin{cases} \Delta \mathbf{W}_{kp} = \partial E_p^{(2)} y_p \left(1 - y_p\right) z_k & (p = 1, 2, 3, \cdots, o) \\ \mathbf{W}_{kp} += \Delta \mathbf{W}_{kp} & (k = 1, 2, 3, \cdots, m) \end{cases} \tag{14}$$

$$\begin{cases} \Delta \mathbf{W}_{ik} = \partial E_k^{(1)} z_k \left(1 - z_k\right) x_i & (k = 1, 2, 3, \cdots, m) \\ \mathbf{W}_{ik} += \Delta \mathbf{W}_{ik} & (i = 1, 2, 3, \cdots, n) \end{cases} \tag{15}$$

By adjusting the connection strengths between the input nodes and hidden layer nodes as well as the connection strengths and thresholds between the hidden layer nodes and output nodes, the error is reduced along the gradient direction. After iterative learning and training, the network parameters (weights and thresholds) corresponding to the minimum error were determined, and training was stopped. The overall architecture of the diagnostic model is shown in Fig. 3. The flowchart of the SDE-BP algorithm is presented in Fig. 4.

In Fig. 4, the initial weights and thresholds of the BP neural network are optimized using the differential evolution algorithm on the right side to improve the training efficiency and performance of the network. During the iteration process, the differential evolution algorithm constantly adjusts the parameters by simulating the evolution process to search for the optimal network structure or parameter combination,

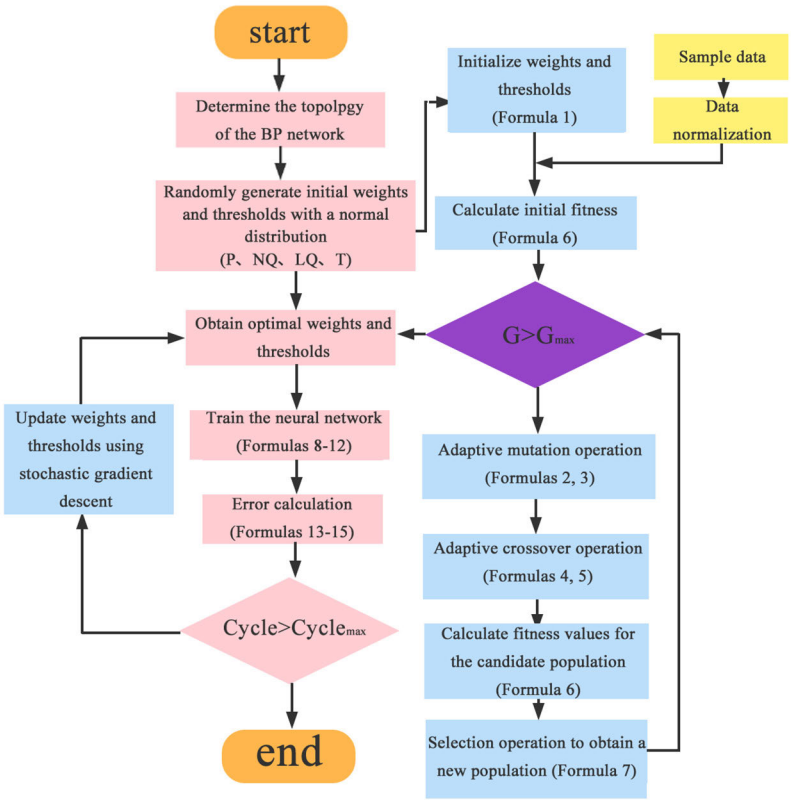


**FIGURE 4.** Flowchart of SDE-BP algorithm.

thereby enhancing the training process of the BP neural network. Meanwhile, during the training process of the BP network, the weights and thresholds were updated through stochastic gradient descent in each round of calculation, gradually converging the iteration to the optimal result. This algorithm uses the maximum number of iterations and loop times as termination conditions to prevent iteration error from being used as a termination condition in different systems, which may result in inappropriate errors and cause a large amount of computation in the system.

### C. INTELLIGENT DECISION-MAKING FOR WATER INJECTION SYSTEM

Most of the research on system decision-making focuses on the optimization of water supply pipeline networks [24], [25], [26], with less involvement in fault decision-making for water injection systems. Given the large amount of production data and incomplete data types in oilfield water injection pipeline network systems, CART decision trees were adopted for fault decision-making. CART decision trees can quickly process large amounts of data, generate viable models in a short time, and represent them graphically, making them easy to understand and interpret [27].

#### 1) DECISION TREES ALGORITHM

Back Propagation (BP) neural networks have been widely used in various types of fault diagnosis [16], but they are rarely used in the fault diagnosis of oilfield water injection systems. The main reason is that the traditional BP neural network has slow convergence owing to the large amount of training data, and it is prone to overfitting, which leads to a local optimum. Optimization of BP neural networks can effectively solve these issues [18], [19]. The proposed SDE-BP algorithm is improved in two ways.

The CART decision trees use the Gini coefficient to determine splits, and the purity of the dataset (D) is represented by the Gini coefficient as follows [28]:

$$\text{Gini}\,(D) = \sum_{i=1}^{n}\sum_{i'\neq i} p_k p_k' = 1 - \sum_{i=1}^{n} p_k^2 \tag{16}$$

The Gini coefficient reflects the probability that two samples randomly drawn from the dataset (D) have inconsistent class labels. The smaller the Gini coefficient (Gini(D) the higher the purity of the dataset (D). The Gini coefficient for discrete data (a) is defined as:

$$\text{Gini}\,(D, a) = \sum_{v=1}^{V} \frac{|D^v|}{|D|}\text{Gini}\,(D^v) \tag{17}$$

Select the attribute with the minimum Gini coefficient as the optimal splitting attribute, that is, Gini_min = min (Gini(D, a)).

#### 2) FAULT DECISION-MAKING

To generate decisions, it is necessary to first construct a decision knowledge system based on multiparameter indicators of

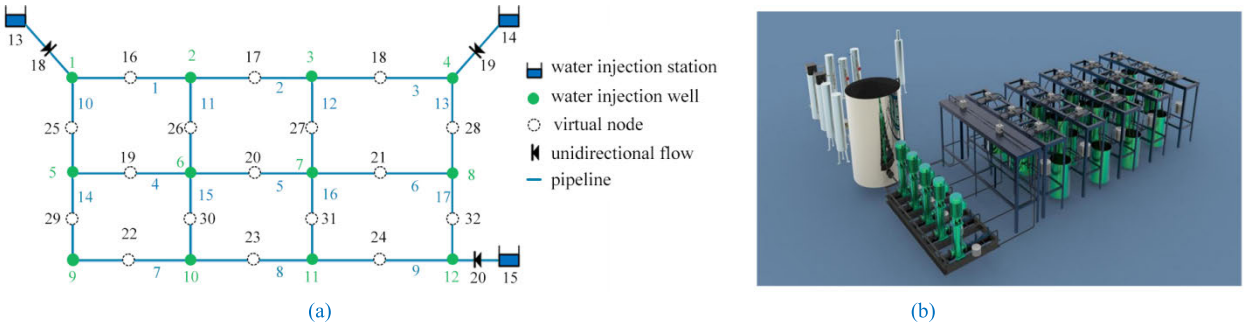


**FIGURE 5.** The simulated water injection network: (a) Structure diagram of the simulated water injection network; (b) Three-dimensional model of the simulated water injection network.

fault points and clarify the fault causes and the classification criteria for decision types. Based on the fault decision-making knowledge system, the CART decision tree algorithm was used to calculate and generate a decision tree using the principle of minimizing the Gini coefficient. The steps are as follows:

- Establish the fault decision-making knowledge system.
- Determine the causes and corresponding measures for a specific fault and perform category classification.
- The CART algorithm was used to calculate the Gini values for various parameters of the training data. The minimum Gini value is selected as the basis for decision tree partitioning and a decision tree for a specific fault is generated.
- When there are multiple categories for a fault, pre-pruning is performed on the decision tree.
- Analysis of the decision tree. Classify the data according to the decision tree classification and match the corresponding causes and solutions for the fault.

## III. EXAMPLE ANALYSIS

Because an actual water injection system cannot be shut down, it is difficult to obtain data or conduct experiments for various types of problems. In this study, EPANET was used to build a water injection network model to simulate the system fault conditions and conduct experimental research on the algorithm.

### A. SIMULATING THE STRUCTURE OF THE WATER INJECTION NETWORK

The simulated structure and three-dimensional model of the water injection network are shown in Fig. 5. The system included 3 water injection stations, 12 water injection wells, and 17 water injection pipelines. The length of the pipelines was 1000 m. The diameter of the pipelines connected to the stations was 500 mm and the diameter of the other pipelines was 300 mm. The friction coefficient was 0.013 and the flow rate of each injection well was 50 $m^3/s$.

### B. FAULT DATA GENERATION

EPANET is a hydraulic software used to simulate and analyze urban water supply and drainage systems. It can simulate flow and pressure distributions in s hydraulic pipeline network. By constructing hydraulic models and calculations, it is possible to evaluate the performance of hydraulic pipeline networks, detect pipeline leaks and blockages, optimize pipeline design and operation, and predict water quality changes [29], [30].

In this paper, the principles for simulating fault conditions are as follows:

- Water Injection Station Faults: Increase, decrease, or stop the pressure at a specific station node. The pressures and flow rates of other injection stations, wells, and pipelines were calculated. Simulate fault conditions such as pump overload, underload, pump shutdown, and low pump outlet flow.
- Water Injection Well Faults: Slightly increasing or decreasing the flow rate at a well node or drastically reducing the flow rate. The pressure and flow rates for the various components were calculated. Simulate fault conditions, such as over-injection, under-injection, valve non-opening, and nozzle clogging in the distributor.
- Pipeline faults: a new node is added at the location of a pipeline leak. The pipeline is divided into two sections. If there is a flow at the node, network leakage is simulated. If adjacent pipelines have no flow or high velocity, pipeline blockage can be simulated. Simulate faults such as blockages, scaling, pipeline bursts, and jetting.

By simulating various types of fault conditions and conducting calculations, 1536 sets of fault data were obtained, with 101 data points in each set. Among these, there were 144 sets of station data, 576 sets of well data, and 816 sets of pipeline data. Each data contained complete parameters, such as pressure and flow at each node of the entire pipeline network, flow of each pipeline segment, and target values. These data served as foundational data for the fault diagnosis model.

### C. DETERMINATION OF PARAMETERS

In the process of training and testing diagnostic models, the setting of parameter values has an impact on the convergence speed and accuracy of the model. To find the optimal hyperparameters, initial settings were set based on the parameter

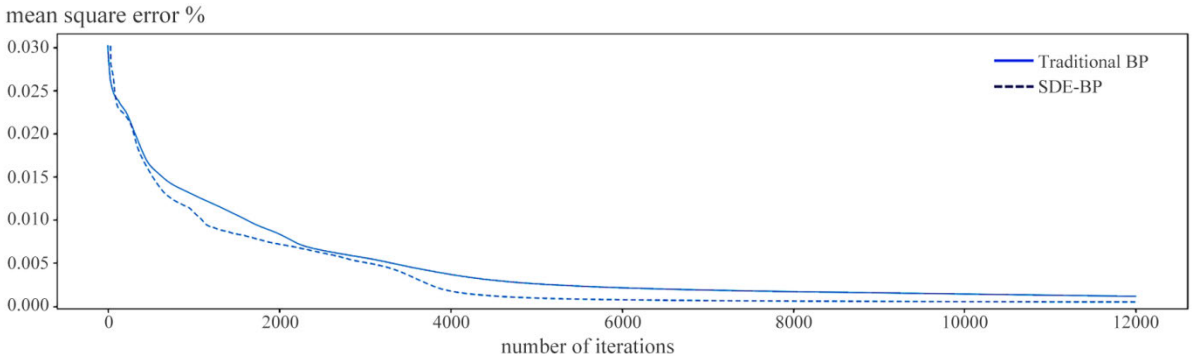


**FIGURE 6.** Mean squared error curve for fault location diagnosis.

values of similar models in the same field. Second, the analogy method is employed to experimentally calculate each parameter individually and to compare and analyze the degree of influence of each parameter to determine the hyperparameters that are suitable for the current state of the pipeline network and have both ideal accuracy and speed. Taking the learning rate as an example, when the learning rate was set to 0.1, the accuracy of the model test was slightly lower (approximately 93%), but still relatively ideal. Although a smaller learning rate can achieve a better accuracy, the convergence of the function becomes slower, resulting in a slower model training speed. Therefore, considering the accuracy and iteration time of the model comprehensively, the learning rate is set to 0.01 to strike a balance between a high recognition rate and efficiency. The settings of the other parameters were also determined using this method.

### D. FIRST-LEVEL OPTIMIZATION NETWORK DIAGNOSTIC ANALYSIS

By simulating all the fault data mentioned above, 90% of the data were selected as the training set for the SDE-BP network model, whereas the remaining 10% were used as the test set. The input layer-hidden layer-output layer structure of the first-level SDE-BP diagnostic model is a three-layer network with 101-20-32 nodes, where the input layer data correspond to the pressure and flow of each node, and the flow data of each pipeline segment. The number of nodes in the output layer is the sum of all stations, wells,

**TABLE 2.** Partial diagnostic data for fault points using an adaptive differential optimized BP network model.

| Fault point | | Station 13 | Station 14 | Station 15 | Well 1 | Well 2 | Well 3 | Well 4 | Well 5 | Well 6 | Well 7 | Well 8 | Well 9 | Well 10 | Well 11 | Well 12 | Pipe 1 | Pipe 2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SDE-BP | sample 1 | 0.012 | 0.023 | **0.971** | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.01 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 |
| | sample 2 | 0.01 | 0.01 | 0.01 | 0.01 | 0.009 | 0.01 | 0.01 | **0.99** | 0.009 | 0.01 | 0.01 | 0.01 | 0.01 | 0.009 | 0.01 | 0.011 | 0.01 |
| | sample 3 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| | sample 4 | 0.07 | 0.012 | 0.01 | 0.006 | 0.012 | 0.013 | 0.011 | 0.011 | 0.012 | 0.07 | 0.012 | 0.015 | 0.009 | 0.014 | 0.007 | 0.010 | 0.008 |
| Traditional BP | sample 1 | 0.023 | 0.018 | **0.950** | 0.015 | 0.009 | 0.007 | 0.014 | 0.011 | 0.009 | 0.011 | 0.015 | 0.014 | 0.010 | 0.010 | 0.024 | 0.012 | 0.019 |
| | sample 2 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | **0.97** | 0.01 | 0.01 | 0.01 | 0.01 | 0.009 | 0.01 | 0.01 | 0.01 | 0.01 |
| | sample 3 | 0.007 | 0.011 | 0.011 | 0.011 | 0.011 | 0.009 | 0.056 | 0.011 | 0.008 | 0.009 | 0.008 | 0.008 | 0.008 | 0.013 | 0.011 | 0.011 | 0.019 |
| | sample 4 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |

| Fault point | | Pipe 3 | Pipe 4 | Pipe 5 | Pipe 6 | Pipe 7 | Pipe 8 | Pipe 9 | Pipe 10 | Pipe 11 | Pipe 12 | Pipe 13 | Pipe 14 | Pipe 15 | Pipe 16 | Pipe 17 | diagnostic results |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SDE-BP | sample 1 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.009 | 0.01 | 0.009 | 0.009 | Station 15 |
| | sample 2 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | Well 5 |
| | sample 3 | 0.01 | 0.01 | **0.986** | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | Pipe 5 |
| | sample 4 | 0.008 | 0.01 | 0.012 | 0.009 | 0.008 | **0.991** | 0.008 | 0.008 | 0.014 | 0.007 | 0.01 | 0.005 | 0.012 | 0.008 | 0.014 | Pipe 8 |
| Traditional BP | sample 1 | 0.006 | 0.008 | 0.009 | 0.016 | 0.012 | 0.008 | 0.006 | 0.011 | 0.011 | 0.01 | 0.008 | 0.088 | 0.022 | 0.01 | 0.006 | Station 15 |
| | sample 2 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | Well 5 |
| | sample 3 | 0.009 | 0.01 | **0.986** | 0.009 | 0.009 | 0.012 | 0.01 | 0.01 | 0.009 | 0.011 | 0.015 | 0.015 | 0.007 | 0.011 | 0.01 | Pipe 5 |
| | sample 4 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | **0.969** | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | Pipe 8 |

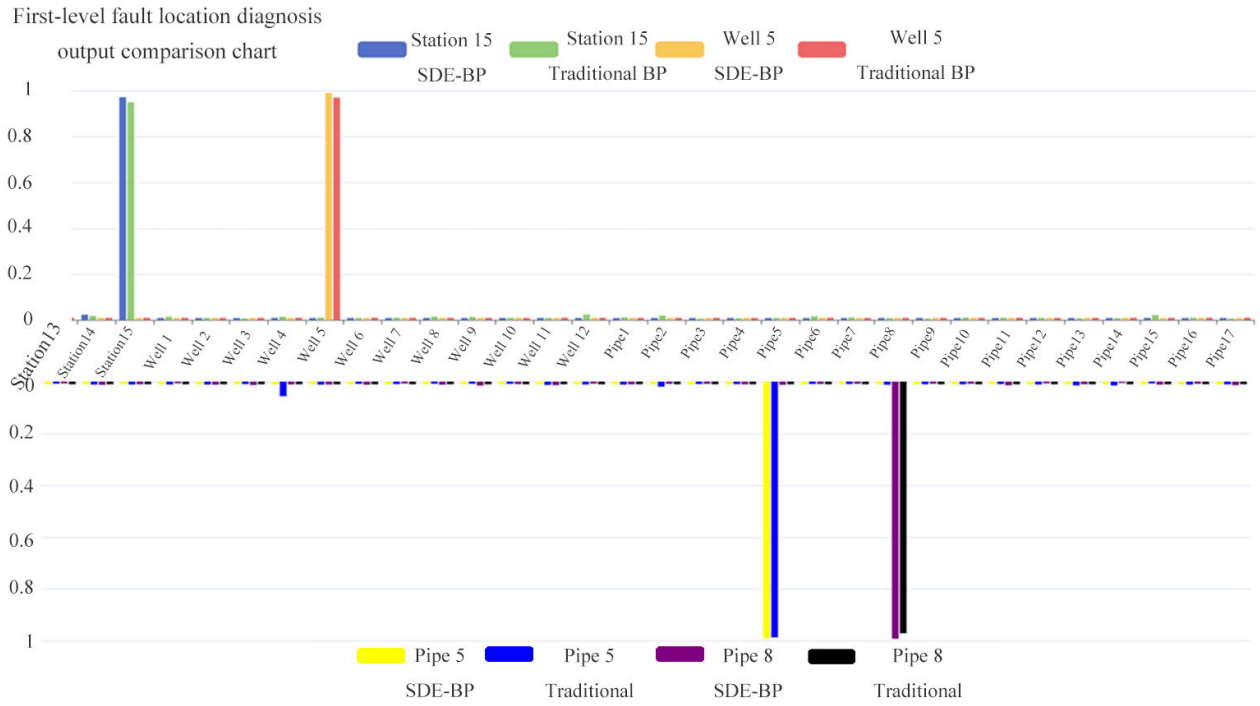


**FIGURE 7.** Comparison chart of first-level fault location diagnosis output results.

and pipeline segments. The maximum number of iterations was set to 12000, the learning rate is 0.01, and the convergence accuracy is 0.00037. The parameters for the adaptive differential evolution algorithm were configured with initial population size of 20, maximum mutation factor of 0.7, a minimum mutation factor of 0.3, maximum crossover factor of 0.9, minimum crossover factor of 0.1, and maximum evolution generation of 50.

Fault location and fault type were diagnosed using both the SDE-BP network model and traditional BP neural network. The mean square error curve for fault location diagnosis with the first-level SDE-BP network is shown in Fig. 6. As can be seen from the figure, the SDE-BP network model demonstrates a tendency to converge after 6000 iterations, achieving a mean square error accuracy of 0.00062. by contrast, the BP network model exhibited a mean square error accuracy of 0.00213 for the same number of iterations. At the maximum number of iterations, the mean square error of the BP network model was 0.00096, which was higher than that achieved by the SDE-BP network model under the same number of training iterations. Comparative analysis revealed that the convergence speed of the SDE-BP network increased by 50% compared with that of the BP network, while its convergence accuracy was enhanced by 56.3%.

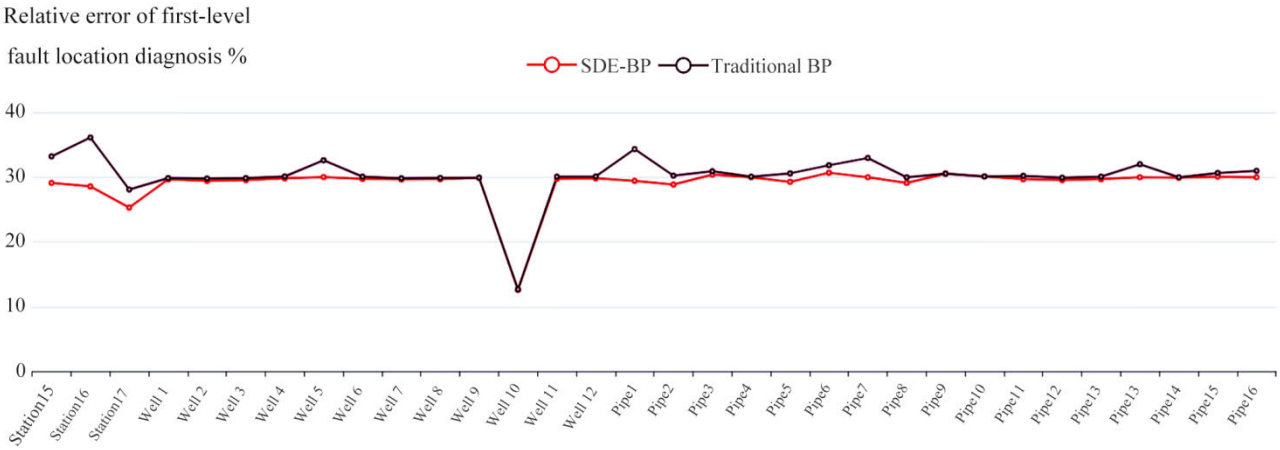


**FIGURE 8.** Relative error chart for first-level fault location diagnosis.

The SDE-BP network model has a three-layer structure with 101, 20, and 32 nodes in each layer. The total number of nodes is not large. Taking the current initial population size of 20 as an example, with a learning rate of 0.01, the simulation on a regular student's work computer had an iteration time of approximately 15.5 seconds. Therefore, the complexity of model training was not high. The computational complexity of the subsequent secondary optimization network diagnosis is similar.

Some of the diagnostic results from the SDE-BP network are presented in Table 2. For brevity, the term "pipeline" in the following Tables and Figures is written as "pipe". As indicated by the bold data in the table, both the first-level SDE-BP model and the traditional BP model are capable of correctly diagnosing and locating all faults in Station 15, Well 5, Pipeline 5 and Pipeline 8. Compared to the BP model, the SDE-BP model produced more accurate output results, leading to an improvement in the diagnostic accuracy of approximately 2.3%. A comparison of the diagnostic output results is shown in Fig. 7.

Fig. 8 depicts the relative error graph for the diagnostic results of various fault points at the first level. As shown in the graph, the relative error for fault location diagnosis using the SDE-BP network model is approximately 30%, which is generally lower than the error rate of the BP neural network model. The fault point of Well 10 exhibits a diagnostic accuracy of 12.672% for both the SDE-BP model and the traditional BP network model, indicating a higher precision.

This verifies that in the first-level fault diagnosis, the SDE-BP network model has a faster training speed and higher training accuracy and can accurately diagnose the location of the pipeline network fault.

### E. SECOND-LEVEL OPTIMIZATION NETWORK DIAGNOSIS

Based on the first-level diagnosis results, diagnostic models can be established for stations, wells, and pipelines. The second-level optimization network topology was set as 101-20-2, with a learning rate of 0.01, and a maximum iteration count of 18000. The parameters for the adaptive differential evolution algorithm were set as follows: initial population size 20, maximum mutation factor 0.7, minimum

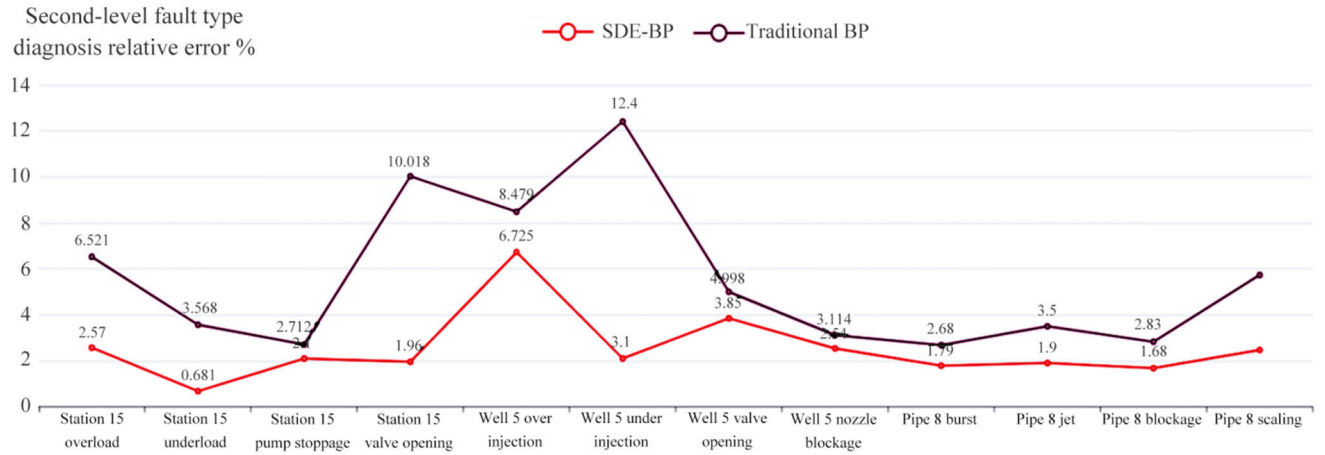


**FIGURE 9.** Relative error chart for second-level fault type diagnosis.

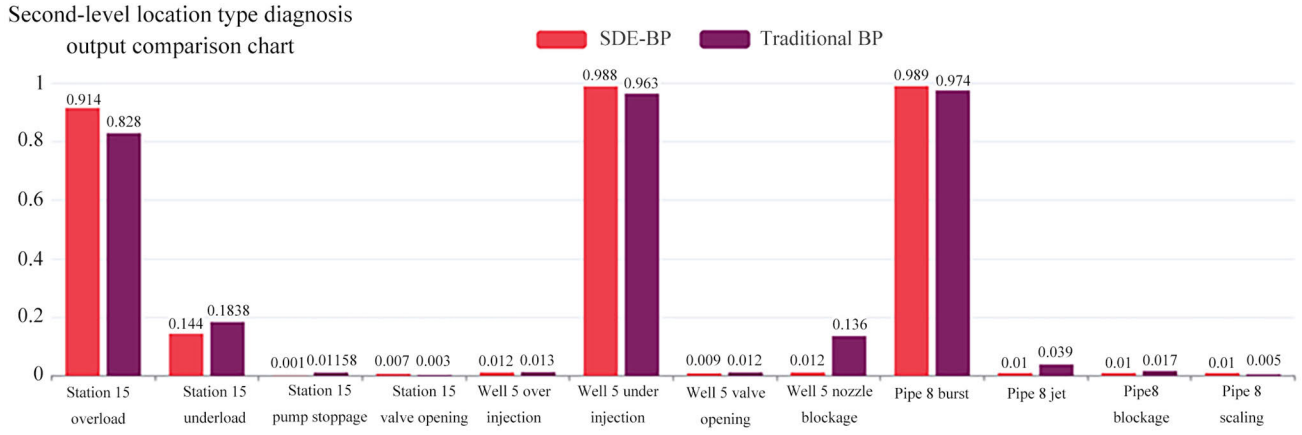


**FIGURE 10.** Comparison chart of fault type diagnostic output results.

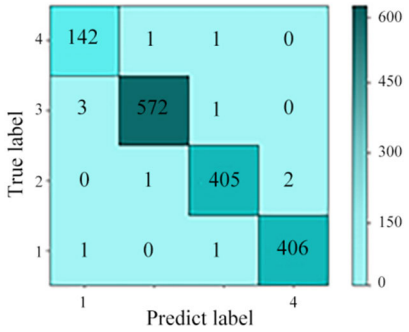


**FIGURE 11.** Confusion matrix for fault type diagnosis.

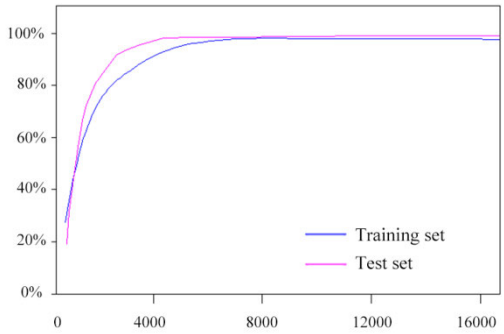


**FIGURE 12.** Changes in accuracy of the diagnostic mode.

mutation factor 0.3, maximum crossover factor 0.9, minimum crossover factor 0.3, and the evolution generations 50. Simultaneously, 32 models with the same structure were established, and second-level diagnosis was performed based on the first-level diagnosis results.

The SDE-BP network model tended to converge after 7000 iterations, with a mean squared error accuracy of 0.0132. The mean squared error accuracy of the traditional BP network model was 0.0419. At the maximum number of iterations, the mean squared error of the SDE-BP model was 0.0086, which represents a 43% improvement in convergence speed and a 47.8% improvement in convergence accuracy compared to the traditional BP network.

For Stations 15, Well 5 and Pipeline 8, a second-level SDE-BP network fault-type diagnosis was conducted, and some error curves are shown in Fig. 9. It can be observed that compared to the SDE-BP network model, the diagnostic error of the traditional BP model is relatively larger. For example, the relative errors of the valve opening of Station 15 and under-injection of Well 5 using the traditional BP model were 10.018% and 12.4%, respectively. In contrast, the relative errors using the SDE-BP model are 1.96% and 2.1%, respectively, which are significantly lower than those of the traditional BP network. For the other parts, the relative errors of the BP model ranged from 0% to 10%, while those of the SDE-BP model ranged from 0% to 7%. For Station 15, Well 5

**TABLE 3.** Diagnostic evaluation indicators for various operating conditions.

| working conditions | precision | recall | F1-score |
|---|---|---|---|
| 1 | 98.63% | 98.61% | 0.962 |
| 2 | 99.65% | 99.30% | 0.990 |
| 3 | 99.26% | 99.26% | 0.986 |
| 4 | 99.50% | 99.50% | 0.987 |

**TABLE 4.** Fault classification and decision-making for water injection stations.

| Fault Type | | Flow | Pressure | Flow/ Pressure | Pipe Flow | Fault Cause | Decision Treatment | Decision Classification |
|---|---|---|---|---|---|---|---|---|
| water injection station | pump overload | ↑≥30% | ↑≥10% | 3.2~4.2 | Equal or not | | | |
| | | Yes | Yes | Yes | Yes | Large pump outlet flow | Replace pump model | 1 |
| | | No | Yes | Yes | Yes | Pump seal part stuck | Pump seal maintenance | 2 |
| | | No | No | Yes | Yes | Pump bearing damage | Pump Bearing replacement | 3 |
| | pump under-load | ↓≥30% | ↓≥10% | 2.5~4.0 | Equal or not | | | |
| | | Yes | Yes | Yes | Yes | Gas in pump body or trapped air in inlet pipe | Fill the pump with water before starting | 1 |
| | | Yes | No | No | Yes | Inadequate closure of pump bottom valve, insufficient water filling | Check the tightness of all valves and leakage of joints of pump | 2 |
| | stop the pump | ↓≥80% | ↓≥60% | 0~2.5 | Equal or not | | | |
| | | Yes | Yes | Yes | No | Reliability of power supply | Check for poor switch contact and blown fuses | 1 |
| | | Yes | No | No | Yes | Pump impeller stuck and blocked | Open pump to remove foreign matter | 2 |
| | | Yes | Yes | No | Yes | Pump bearing bending | Replace parts | 3 |

**TABLE 5.** Fault classification and decision-making for water injection wells.

| Fault Type | | Flow | Pressure | Flow/ Pressure | Pipe Flow | Fault Cause | Decision Treatment | Decision Classification |
|---|---|---|---|---|---|---|---|---|
| water injection well | over-injection | ↑≥20% | ↑≥5% | 1.2~1.5 | Equal or not | | | |
| | | Yes | Yes | No | Yes | Overload of water injection station | Replace pump model | 1 |
| | | Yes | Yes | Yes | Yes | Valve falling off | Check valve for damage, regular maintenance and replacement | 2 |
| | | Yes | No | Yes | Yes | Packer failure | Improve tool quality, reduce wellbore harm and eliminate pipe string creep | 3 |
| | | No | No | No | Yes | Metering device abnormal | Check the water meter, repair and replace it in time | 4 |
| | under-injection | ↓≥20% | ↓≥5% | 0.7~1.2 | Equal or not | | | |
| | | Yes | Yes | Yes | Yes | Pipe network perforation damage | Strengthen pipeline corrosion prevention and replace regularly according to life | 1 |
| | | No | Yes | Yes | Yes | Scaling and clogging of surface pipeline | Regularly clean surface pipelines to reduce impurities deposited in pipes | 2 |
| | | Yes | No | No | Yes | Pump stop at station | Enhance coordination of stations and fewer pump stoppages | 3 |

**TABLE 6.** Fault classification and decision-making for water injection pipelines.

| Fault Type | | Flow | Pressure | Flow/ Pressure | Pipe Flow | Fault Cause | Decision Treatment | Decision Classification |
|---|---|---|---|---|---|---|---|---|
| water injection pipeline | tube explosion | ↓≥50% | ↓≥10% | 0.7~1.5 | Equal or not | | | |
| | | Yes | Yes | Yes | No | Pipe material life | Replace ductile iron and steel pipes | 1 |
| | | No | No | Yes | No | Temperature changes | Install pipeline expansion joints reasonably | 2 |
| | | No | Yes | Yes | No | The water hammer effect | Reduce the closing speed of the valve | 3 |
| | | Yes | No | Yes | No | Construction and installation | Strictly control the laying and installation of pipelines | 4 |
| | jet | ↓≥10% | ↓≥5% | 1.3~1.5 | Equal or not | | | |
| | | Yes | No | Yes | No | Pipeline interfaces subjected to lateral forces | Improve the tensile strength of the pipeline material | 1 |
| | | Yes | No | No | No | Pipeline corrosion | Pipeline anti-corrosion treatment | 2 |
| | blockage | ↓≥50% | ↑≥50% | 0.4~1.4 | Equal or not | | | |
| | | Yes | No | Yes | No | Filter blockage | Timely cleaning of filters | 1 |
| | | Yes | Yes | Yes | No | Foreign object blockage | Disassembly and cleaning of pipelines | 2 |

and Pipeline 8, the average relative errors in the fault category diagnosis using the SDE-BP model and the traditional BP model were 2.381% and 5.545%, respectively. The optimized average relative error was reduced by 3.164%.

A comparative diagram of the partial secondary fault-type diagnosis results is presented in Fig. 10.

This indicated an improvement in the diagnostic accuracy of nearly 10%. Additionally, the SDE-BP network correctly diagnosed all fault types, including pump startup at Station 15, over-injection of Well 5, leakage of Pipeline 8, blockage of Pipe 12, and other fault types, with a perfect accuracy rate of 99%. It can be concluded that the optimized BP neural network with the adaptive differential evolution method can enhance diagnostic accuracy.

By comparing the results with the true labels, the confusion matrix under the four working conditions of station failure, well failure, pipeline leakage and scaling blockage is shown in Fig. 11. Table 3 presents the evaluation metrics for each

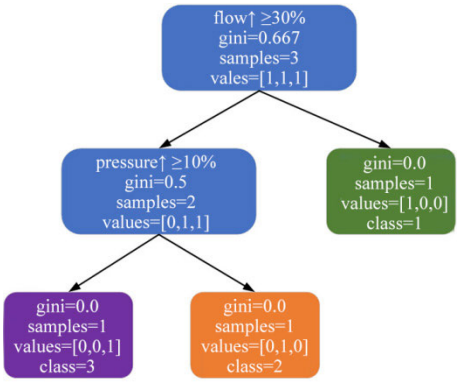


**FIGURE 13.** Decision tree for water injection station 15.

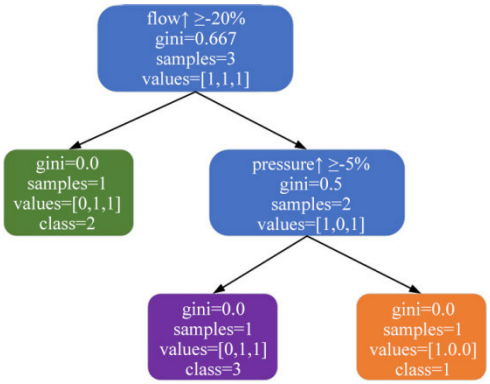


**FIGURE 14.** Decision tree for water injection well 5.

working condition. The accuracy rates during the training and testing phases are shown in Fig. 12.

### F. DECISION GENERATION

In this study,a multiclassification decision tree, which is generated based on the CART algorithm, is adopted as the decision model. Based on the variation ranges of four indicators, including flow rate, pressure, pipeline flow rate and flow-pressure ratio, decision categories are classified. Three decision tree classification tables for water injection stations, wells, and pipelines were established, as shown in Tables 4, 5, and 6.

By combining the fault decision tables, decision trees are generated based on the fault types determined from the second-level diagnosis of the SDE-BP fault diagnosis model. During the parameter discrimination process, conditions are represented by 1 if satisfied and 0 if not the satisfied, from which intuitive strategies can be adopted when a specific fault occurs. Python was used to write a decision tree algorithm to improve water injection efficiency. Taking Station 15 and Well 5 as examples, the decision can be generated as follows:

Fig. 13 represents the decision tree for the pump overload fault type of Station 15. The three decision-making criteria are as follows: first, calculate the current minimum Gini coefficient, which is 0.667, and check whether the flow rate of Station 15 exceeds 30% of the normal injection amount be divided into Class 1. At this point, the Gini coefficient decreased to 0.5. Second, check whether the fault pressure exceeds 10% of the normal pressure to be divided into Classes 2 and 3. The Gini coefficient became zero, indicating that the decision-making process was complete. According to the decision tree, when a water injection station is overloaded, decisions can be made based on the node flow rate, pressure, flow rate, pressure ratio, and flow parameters at both ends of the pipeline during failure. If it falls into Class 1, it suggests the pump model should be changed. If it falls into Class 2, pump seal maintenance is required. In this example, Station 15 falls into Class 3. The replacement of the pump bearings is recommended.

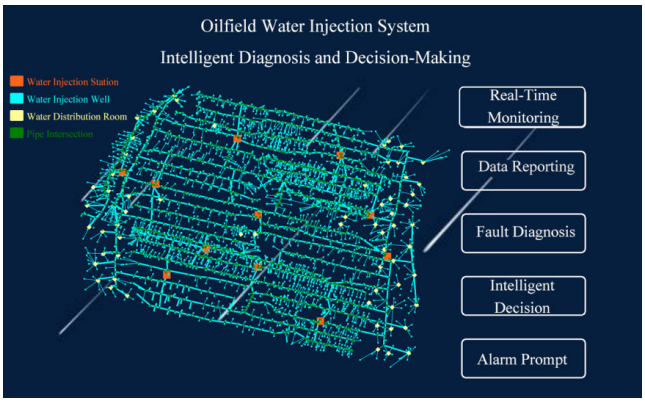


**FIGURE 15.** Intelligent diagnosis and decision-making platform.

Fig. 14 shows the decision tree for the fault of the under-injection of Well 5. The three decision-making criteria are as follows: First, the current minimum Gini of 0.667 is calculated. We need to check the flow rate of Well 5, if it is below 20% of the normal injection amount, the fault can be classified as Class 2. At this point, the Gini coefficient decreased to 0.5. Second, it needs to be checked whether the fault pressure is below 5% of the normal pressure to be divided to the Classes 1 and 3. The Gini coefficient was 0. The decision-making division was completed. In this example, water injection Well 5 falls into Class 2. This decision suggests the need for pipeline scaling and periodic cleaning. If it falls into Class 1, the strategy will Strengthen anti-corrosion and periodic replacement. If it falls into Class 3, the strategy will strengthen the coordination of the water injection stations.

### G. DEVELOPMENT OF INTELLIGENT DIAGNOSIS AND DECISION-MAKING PLATFORM

In this study, an intelligent diagnosis and decision-making platform for a water injection pipeline network was constructed by combining Python and JavaScript languages and utilizing tools such as Flask, PyQt5, CSS, ECharts, and MySQL [31], [32], [33], [34]. This platform is depicted in Fig. 15, and its interface incorporates functions such as

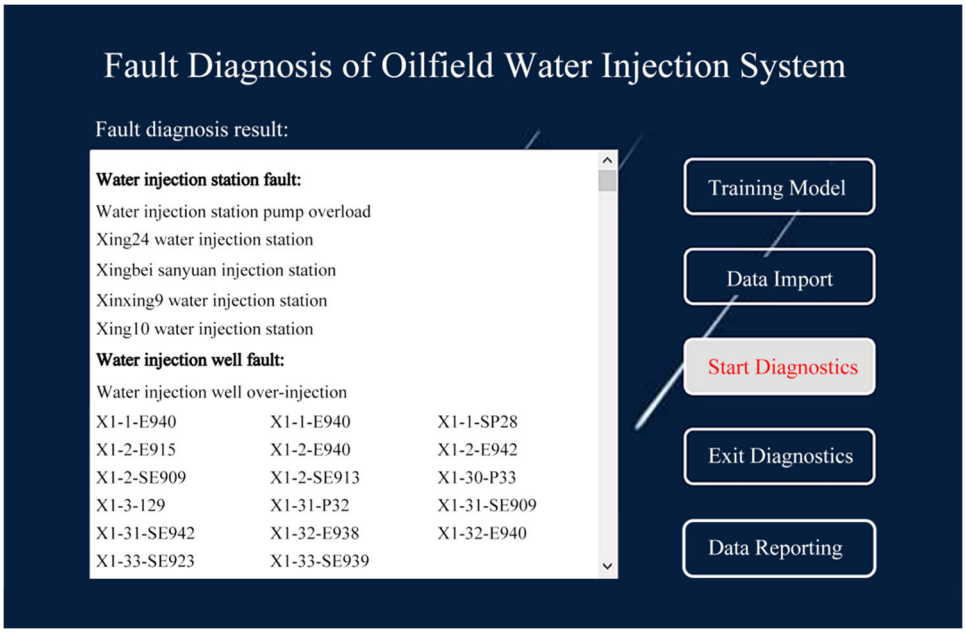


**FIGURE 16.** Fault diagnosis and diagnostic results.

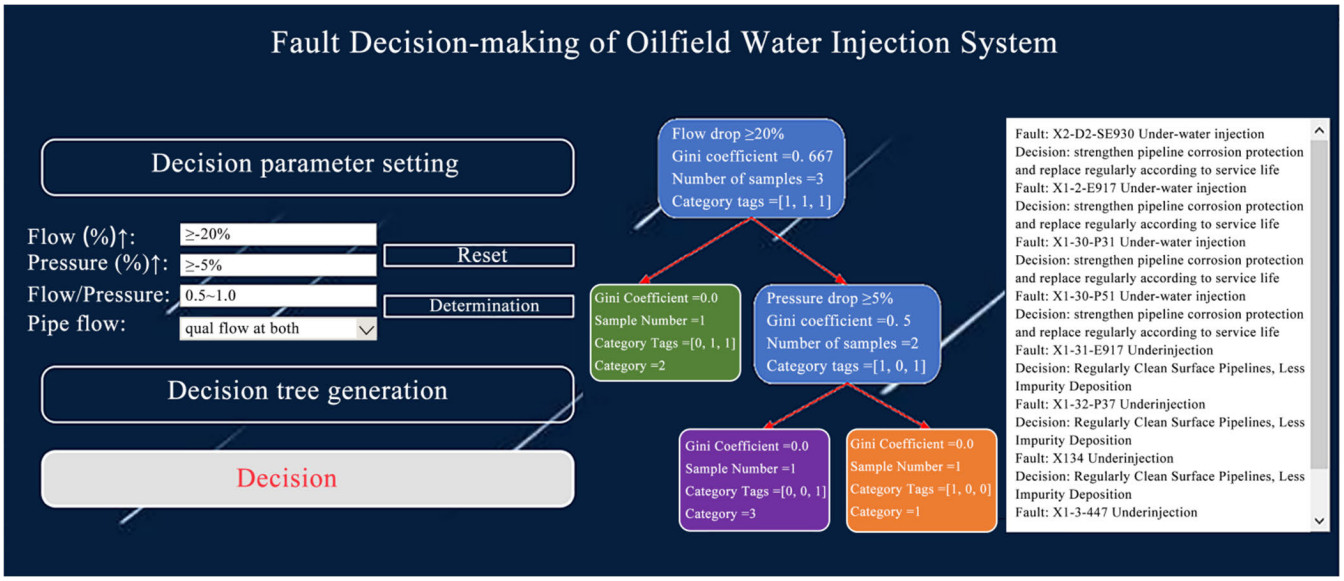


**FIGURE 17.** Oilfield water injection system fault decision.

real-time monitoring, data reporting, fault diagnosis, intelligent decision-making, and alarm prompts.

The production data of the water injection system shown in Fig. 1 were used to perform the fault diagnosis and decisions of the system, as shown in Fig. 16 and 17.

This study is the first to apply the SDE-BP model and CART decision tree to research faults in oilfield water injection systems, achieving good results. However, in the diagnosis of the primary fault locations, there are some locations with slightly lower prediction accuracy. To address this, the number of nodes in the input data of the model can be increased to establish a more detailed model. For more complex fault issues, such as large-scale looped pipeline networks and multistation pipeline networks with multiple faults, further experimental verification is required.

## IV. CONCLUSION

This study aims to establish an intelligent oilfield, and focuses on fault diagnosis and decision-making methods for large and complex water injection systems. The following conclusions were drawn.

Regarding faults in the oilfield water injection system, we innovatively established various fault diagnosis criteria and fault trees. We categorized the faults into three main areas: water injection stations, water injection wells, and pipelines. Within these areas, we developed a comprehensive fault classification system that includes 12 subcategories, encompassing more than 20 specific issues, along with the corresponding diagnostic criteria. Furthermore, we propose a method for diagnosing faults in a pipeline network by utilizing pressure and flow parameters.

For the first time, the method of optimizing a BP neural network using a self-adaptive differential evolution algorithm (SDE-BP) was employed to construct a two-level fault diagnosis model for water injection systems, achieving a comprehensive diagnosis of fault locations and types. Compared with the traditional BP algorithm, the proposed algorithm demonstrated better diagnostic accuracy and faster convergence speed. In the first-level fault location diagnosis, the convergence speed increased by 50%, the accuracy improved by 56.3%, and the diagnostic accuracy rate increased by nearly 2.3%. In the second-level fault-type diagnosis, the SDE-BP model also exhibited superior diagnostic performance, with a 43% increase in convergence speed and a 47.8% improvement in convergence accuracy. In the diagnosis of station valve openings, the relative error was reduced from 10.018% to 1.96%, and the error in the under-injection of wells was reduced from 12.4% to 2.1%. In the diagnosis of pipeline nodes, the relative error improved by at least 3%. The mean relative error of the SDE-BP model was reduced by 3.164%. From the experiment, the SDE-BP model has an ideal effect on the diagnosis of water injection pipeline network failure and can provide a reference for pipeline networks of petroleum, natural gas, water supply, and heating supply.

Based on the fault diagnosis results of the SDE-BP model, we established a fault decision-making knowledge system for water injection stations, wells, and pipeline networks. The decision categories were divided based on the changes in four indicators: flow rate, pressure, pipeline flow rate, and flow-pressure ratio. The CART decision tree algorithm was used to generate decisions, with the algorithm being coded in Python, ultimately providing an intuitive strategy for handling a specific fault when it occurs.

An intelligent diagnosis and decision-making platform for water injection systems has been established. Using this platform, fault diagnosis and decision-making were conducted on a specific water injection system in the Daqing Oilfield in China. This study can enhance the scientific management of oilfields and promote the development of smart oilfields.

## REFERENCES

[1] E. Q. Zhan, "Issues and energy-saving analysis of oilfield water injection pipeline network," *Chem. Eng. Manage.*, vol. 568, no. 25, pp. 195–196, Sep. 2020.

[2] P. Guo, K. Wang, and A. L. Luo, "Research status and prospects of computational intelligence in big data analysis," *J. Softw.*, vol. 26, no. 11, pp. 3010–3025, Nov. 2015, doi: 10.13328/j.cnki.jos.004900.

[3] S. L. Han, X. J. Liu, and D. D. Wang, "On the development and future direction of water injection," *China New Technol. New Products*, vol. 20, no. 8, p. 135, Apr. 2013, doi: 10.13612/j.cnki.cntp.2013.08.011.

[4] T. Kim, K. Kim, J. Hyung, and J. Koo, "Integrated water suspension risk assessment using fault tree analysis and genetic algorithm in water supply systems," *Desalination Water Treat*, vol. 227, pp. 104–115, Jul. 2021, doi: 10.5004/dwt.2021.27358.

[5] A. Lindhe, T. Norberg, and L. Rosén, "Approximate dynamic fault tree calculations for modelling water supply risks," *Rel. Eng. Syst. Saf.*, vol. 106, pp. 61–71, Oct. 2012, doi: 10.1016/j.ress.2012.05.003.

[6] L. H. Song, "Fault automated diagnosis method for township intelligent water supply systems," *Water Conservancy Sci. Technol. Economy*, vol. 29, no. 2, pp. 152–156, Feb. 2023.

[7] C. W. Qi, X. X. Cao, and Z. H. Qie, "Research on burst pipe diagnosis model of water supply pipeline network based on HHT and SVM," *Water Hydropower Energy Sci.*, vol. 35, no. 2, pp. 153–156, Feb. 2017.

[8] Y. Zhang, Y. Yao, and Z. Chen, "Power system fault diagnosis model via particle swarm differential evolution-based extreme learning machine," *Machinery Electron.*, vol. 42, no. 3, pp. 60–64, Mar. 2024.

[9] W. S. Dong, D. Wang, and Z. N. Wang, "Research on pipeline network fault diagnosis based on improved GAAA algorithm and BP neural network," *District Heating*, no. 5, pp. 9–11, Oct. 2017, doi: 10.16641/j.cnki.cn11-3241/tk.2017.05.003.

[10] L. M. Ji, "Leakage fault diagnosis of centralized heating pipeline network based on deep learning," M.S. thesis, Dept. Thermal Eng., Shandong Jianzhu Univ., Qingdao, China, 2022.

[11] S. M. Liu, B. L. Miao, and Z. D. Lou, "Energy consumption evaluation and auxiliary decision-making technology research and application of oilfield water injection system," *Energy Conservation Petroleum Petrochemicals*, vol. 12, no. 6, pp. 27–30, Jun. 2022.

[12] J. Yan, X. Liu, and J. Pang, "Construction and application of automatic water injection decision analysis system for changqing oilfield water injection station," *China Sci. Technol. Inf.*, vol. 510, no. 1, pp. 134–135, Feb. 2015.

[13] K. Boryczko, D. Szpak, J. Zywiec, and B. Tchórzewska-Cieślak, "The use of a fault tree analysis (FTA) in the operator reliability assessment of the critical infrastructure on the example of water supply system," *Energies*, vol. 15, no. 12, p. 4416, Jun. 2022, doi: 10.3390/en15124416.

[14] J. Li, "Research on visualization, diagnosis, and decision-making techniques for oilfield water injection pipeline network systems," M.S. thesis, Dept. Mech. Sci. Eng., Northeast Petroleum Univ., Daqing, China, 2023.

[15] X. H. Wang, "Research on leakage diagnosis and localization of oilfield water injection pipeline network," M.S. thesis, Dept. Mech. Sci. Eng., Northeast Petroleum Univ., Daqing, China, 2019.

[16] H. W. Zhu, "Development and application of expert system for intelligent diagnosis and decision-making in oilfield water injection system," M.S. thesis, Dept. Mech. Sci. Eng., Northeast Petroleum Univ., Daqing, China, 2022.

[17] X. Zhao and S. Y. Xue, "Development and application of artificial intelligence in fault diagnosis," *Internet Things Technol.*, vol. 12, no. 7, pp. 71–72, Jul. 2022, doi: 10.16667/j.issn.2095-1302.2022.07.021.

[18] P.-H. Zhang, J.-L. Zhang, W.-W. Ren, J. Xie, M. Li, J.-Z. Li, F. Ding, J.-K. Wang, and Z.-R. Dong, "Identification of waterflooded zones and the impact of waterflooding on reservoir properties of the funing formation in the subei basin, China," *J. Zhejiang Univ. Sci. A*, vol. 14, no. 2, pp. 147–154, Feb. 2013.

[19] Y. Jiang, H. Zhang, K. Zhang, J. Wang, J. Han, S. Cui, L. Zhang, H. Zhao, P. Liu, and H. Song, "Waterflooding interwell connectivity characterization and productivity forecast with physical knowledge fusion and model structure transfer," *Water*, vol. 15, no. 2, p. 218, Jan. 2023, doi: 10.3390/w15020218.

[20] D. C. Zou and F. S. Bai, "Neural network optimization design based on differential evolution algorithm," *J. Chongqing Normal Univ.*, vol. 39, no. 1, pp. 79–89, Feb. 2022.

[21] L. Tang, J. Li, W. Lu, P. Lian, H. Wang, H. Jiang, F. Wang, and H. Jia, "Well control optimization of waterflooding oilfield based on deep neural network," *Geofluids*, vol. 2021, pp. 1–15, Apr. 2021, doi: 10.1155/2021/8873782.

[22] W. Zou, Z. B. Wang, and Y. N. Wang, *Deep Learning: Design Examples Based on MATLAB*. Beijing, China: Beihang Univ. Press, 2020, pp. 60–65.

[23] Y. J. Lu, *Introduction To Deep Learning: Theory and Implementation Based on Python*. Beijing, China: Posts & Telecom Press, 2018, pp. 7–28.

[24] D. Morais and A. Almeida, "Water supply system decision making using multicriteria analysis," *Water SA*, vol. 32, no. 2, pp. 229–235, Dec. 2007.
[25] Y. Gao, Y. Li, and T. Y. Song, "A decision support system for water supply emergency management with multiple sources," *J. Water Supply Res. Technol.-Aqua*, vol. 65, no. 2, pp. 135–144, Feb. 2016, doi: 10.2166/aqua.2015.086.
[26] K. Cockerill, "The water supply is fine: Decision-maker perceptions of water quantity and supply-side management," *Water Environ. J.*, vol. 28, no. 2, pp. 242–251, Jun. 2014, doi: 10.1111/wej.12029.
[27] G. Aurelian, *Hands-On Machine Learning With Scikit-Learn, Keras, and TensorFlow*. Beijing, China: Machine Press, 2020, pp. 9–14.
[28] J. X. Wang, "Research and practical application of decision tree algorithm principles," *Comput. Program. Skills Maintenance*, vol. 446, no. 8, pp. 54–56, Aug. 2022, doi: 10.16184/j.cnki.comprg.2022.08.043.
[29] M. Gheibi, N. Emrani, M. Eftekhari, K. Behzadian, M. Mohtasham, and J. Abdollahi, "Assessing the failures in water distribution networks using a combination of geographic information system, EPANET 2, and descriptive statistical analysis: A case study," *Sustain. Water Resour. Manage.*, vol. 8, no. 2, p. 47, Feb. 2022.
[30] I. Iulian, D. A. Nicolae, and G. S. Carmen, "EPANET modeling of an urban groundwater field," in *Proc. ES Web Conf.*, vol. 85, 2019, p. 6003, doi: 10.1051/e3sconf/20198506003.
[31] R. J. Zhang, D. W. Wang, and H. S. Zhang, "Preliminary exploration on data visualization design of oilfield water injection system," *Ind. Des.*, no. 8, pp. 83–84, Aug. 2021.
[32] B. Wei, L. Lu, Q. Li, H. Li, and X. Ning, "Mechanistic study of oil/brine/solid interfacial behaviors during low-salinity waterflooding using visual and quantitative methods," *Energy Fuels*, vol. 31, no. 6, pp. 6615–6624, Jun. 2017.
[33] O. E. Abdelaziem, A. Gawish, and S. F. Farrag, "Application of computer vision in machine learning-based diagnosis of water production mechanisms in oil wells," *SPE J.*, vol. 28, no. 5, pp. 2365–2384, Oct. 2023, doi: 10.2118/211804-pa.
[34] T. Z. Lv, Y. Yan, and Z. W. Liu, "Research data visualization system based on flask and ECharts," *Comput. Telecommun.*, no. 11, pp. 16–19, Nov. 2020, doi: 10.15966/j.cnki.dnydx.2020.11.005.

**RUIJIE ZHANG** received the B.S. degree in petroleum field machinery, the M.S. degree in mechanical design and theory, and the Ph.D. degree in oil and gas field surface engineering from Northeast Petroleum University, in 1997, 2000, and 2011, respectively. She is currently an Associate Professor with Northeast Petroleum University, China. Her research interests include virtual reality, big data visualization, system simulation and optimization, and industrial design. In 2018, she served as a Key Member for the Key Program of International Science and Technology Innovation Cooperation between China and United States Government and went to the University of California at Irvine (UCI) for a short-term project exchange.

**WENTING YANG** received the B.S. degree in product design from Qingdao University of Science and Technology, in 2022. She is currently pursuing the master's degree in industrial design engineering with Northeast Petroleum University. She obtained six national appearance patents.

**JIE LI** received the bachelor's degree from Northeast Petroleum University, where he is currently pursuing the master's degree with the School of Mechanical Science and Engineering. He majored in intelligent optimization of oilfield water injection, referring to artificial intelligence, data visualization, population optimization, and many other studies.

**SHENGLIANG GAO** received the bachelor's degree in mechanical engineering from Lanzhou City College, in 2022. He is currently pursuing the master's degree in mechanical engineering with Northeast Petroleum University. His research field is petroleum machinery, with a focus on the design and performance analysis of oilfield wellhead packing adding devices.

**YAN WANG** received the B.S. degree in chemical machinery and equipment and the M.S. degree in chemical process machinery from Northeast Petroleum University, in 2001 and 2004, respectively, and the Ph.D. degree in petroleum and natural gas engineering from Northeast Petroleum University, in 2013. She was a Visiting Scholar with the University of California at Irvine (UCI) in 2018, and currently is a Professor at Northeast Petroleum University. She published more than 30 chapters in English and Chinese, of which more than ten are in EI and ISTP, and participated in more than 40 projects, such as the Key Program of International Science and Technology Innovation Cooperation between China and United States Government, and the Natural Science Foundation of Heilongjiang Province.

**SHENG GAO** received the B.E. degree in mechanical manufacturing from Jilin University of Technology, in 1992, the M.E. degree in oil and gas mechanical engineering from Northeast Petroleum University, in 1995, and the Ph.D. degree in mechanical and electronic engineering from Harbin Institute of Technology, in 2004. He is currently a Professor, the Doctoral Supervisor, and the Reserve Leader of the Mechanical Design and Theory Discipline of the Provincial Leading Talent Echelon. He is an Editorial Board Member of *China Petroleum Machinery*, a Senior Member of the Chinese Society of Mechanical Engineering, Promoted to an Associate Professor, in 2003, and a Professor, in 2006 (Broken Qualification) presided over and completed a number of topics, such as the National Key Research and Development Program, the National Science and Technology Tackling Program, China Postdoctoral Science Fund, the Provincial Natural Science Fund, the Provincial Science and Technology Tackling Program, the PetroChina Science and Technology Innovation Fund, and the Science and Technology Innovation of Hainan Province.