Research Article

Human Reliability Evaluation Based on Objective and Subjective Comprehensive Method Used for Ergonomic Interface Design

Xin Liu [iD],[1] **Zheng Liu** [iD],[2] **Peng-qing Chen** [iD],[1] **Zi-yu Xie** [iD],[1] **Bin-jun Lai**,[2] **Bo Zhan**,[2] **and Jian-rou Lao**[2]

[1]*Industrial Design Specialty, Guangzhou University, Guangzhou 510006, China*
[2]*School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou 510006, China*

Correspondence should be addressed to Zheng Liu; liu_best@yeah.net

In the process of machining, the operator's incorrect operation using the control human-computer interaction interface may lead to a disastrous effect on the whole production chain such as task failure and system failure. The eight types of using errors about human-computer interaction interface system (misperception, memory lapse, carelessness, improper decision-making, improper operation, improper interface layout design, improper icon and text display design, and improper prompt feedback) are proposed based on the cognition from users. According to the eight types of using error, the human reliability and safety evaluation method is proposed. Considering the variability objective and uncertainty subjective factors in the human-machine interaction process, the human reliability evaluation method is proposed based on objective and subjective comprehensive weighting under ergonomic principles. Firstly, based on the investigation of relevant machine tool interface design, the comprehensive evaluation index system for man-machine reliability of complex interface design is established. Secondly, an objective and subjective comprehensive weight assignment method as the composition of Theil's entropy weight method combined with TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) and improved AHP (analytic hierarchy process) method combined with fuzzy mathematics is proposed. It can not only effectively avoid the impact of excessive subjective factors on empowerment but also fully excavate the real information in the survey data and improve the objectivity and accuracy of the weight data of evaluation indicators at all levels. Thirdly, taking the interface design of CNC machine tool equipment for example, the best human-computer interaction interface design schemes are chosen according to the human reliability evaluation method based on objective and subjective comprehensive method calculated by Entropy-TOPSIS-AHP. Finally, based on eye movement experiments, the use efficiency and accuracy of the human-computer interaction interface design scheme chosen by the assessment method are better than the other design schemes or original one to verify the assessment method validity.

1. Introduction

The research core of ergonomic reliability analysis is based on the analysis, prediction, reduction, and prevention of man errors in addition to the qualitative and quantitative analysis and evaluation carried out for man reliability [1, 2]. In the process of complex system operation, user's errors may cause task failure, system failure, or system crash and

even lead to serious accidents [3, 4]. Nowadays, the research of reliability analysis mainly focuses on the mechanical structure and mechanical experiments [5]. However, more and more attention has been paid to the research of ergonomic reliability. Research on ergonomic reliability has established models and related technologies to study the occurrence mechanism, evaluation, and prevention of faults and accidents, applied them to various methods to improve

human reliability [6], and mainly used them in the medical field [7, 8]. Because of the subjective complexity of human reliability of the working interface system, a few mathematical models of human reliability evaluation are applied on working interface design and there is a lack of quantitative research. Nowadays, the majority of research about interface design is in the field of computing for better programming and technical quality [9], and the ergonomic research about interface design is growing gradually [10], but there is little research on human reliability about interface design and it is focused on the field of nuclear power plant based on probability mapping function mainly [11]. Because of the large amount of information and complex information relationship of the man-machine interaction in the working interface system, unlike the common entertainment product interface, it needs more rigorous cooperation between people and the system, and the organization and presentation of information are more scientific [12]. However, unlike structural reliability analysis [13], human reliability analysis must consider the subjective factors of users. Therefore, it is very necessary to propose a reasonable human reliability assessment method. There are so many human reliability analysis methods and quantification techniques.

The Technique for Human Error Rate Prediction (THERP) is used for a specific action in the process of human cognition [14]. Human Error Assessment and Reduction Technique (HEART) is used to analyze the human error probability [15]. Cognitive Reliability and Error Analysis Method (CREAM) provides a framework of Common Performance Conditions (CPCs) to implement the subjective human error probability estimation from expert judgment based on performance shaping factors [16]. In addition, there are many applications based on entropy and entropy weight method in the area of reliability and safe risk evaluation. Mahdy proposed a new kind of weighted entropy measure which is a new information measure based on its properties according to stochastic orders and reliability theory [17] Sandoval analyzed risk assessment of flooding reliability based on entropy [18]. Entropy theory is also applied in the field of human reliability analysis. El-Ladan made a human reliability analysis based on human entropy boundary conditions applied in the field of marine and offshore [19]. Integration of entropy and TOPSIS can be used to deal with the problem of reliability analysis effectively. Mohammed solved the issues of benchmarking and selection by the COVID-19 diagnosis model established based on entropy and TOPSIS successfully [20]. In addition, entropy combined with AHP applied to evaluate comprehensively is effective and useful. Nagpal applied fuzzy AHP and entropy approach in usability evaluation of website system [21]. AHP is mainly used for subjective evaluation. AHP is often used for human reliability evaluation to select the human factors research method. Petruni applied AHP effectively to choose the appropriate human reliability analysis method in the field of the automotive industry [22]. AHP accompanied by fuzzy sets theory is also often applied in risk and reliability evaluation [23]. Jasra used AHP and fuzzy mathematics to make a reliability assessment of the software system [24]. Although entropy, TOPSIS, and AHP are applied in human reliability and risk evaluations effectively, they are applied in the human reliability for ergonomic interface design rarely, and human reliability of complex working interface is always ignored more easily than the internal structure research.

According to the above problems based on the reliability theory dealing with uncertainties [25], a comprehensive human reliability evaluation index system for working control interface design based on relevant research is established [26, 27]. The ergonomic reliability must balance the quantification of objective and subjective uncertainties [28]. A comprehensive weight assignment method combining subjective and objective is proposed, which can not only effectively avoid the excessive subjective factors to impact empowerment but also fully tap the real information in the survey data and improve the objectivity and accuracy of the weight data of evaluation indexes at all levels of ergonomic reliability [29, 30]. Then, Theil's entropy weight method is applied in the objective evaluation and TOPSIS is provided to make the quantitative evaluation. In addition, the subjective evaluation is proposed based on improved AHP and fuzzy mathematics transformed into quantitative evaluation, and the evaluation results can be analyzed and fed back to guide the design process. Finally, the validity and the effectiveness of the evaluation method proposed in this research is verified by the experiment of the eye tracker.

2. Human Reliability Evaluation Model for Control Interface Design

The evaluation process of human reliability is illustrated in Figure 1. Owning to the subjectivity of human thought and behavior characteristics, the objective and subjective factors should be considered comprehensively for the human reliability evaluation. The entropy weight method is used to determine the objective weights of indicators in addition to the decision matrix based on TOPSIS, while the improved AHP is applied in calculating the subjective weighs of indicators in addition to the decision matrix based on fuzzy mathematics. According to the methods mentioned above, the objective and the subjective human reliability evaluations are built, respectively. Then, eye-tracking experiments are used to verify the usefulness and effectivity of the human reliability evaluation method.

2.1. Evaluation Indexes Selected for Ergonomic Interface Design. Reasonable human reliability evaluation of complex interface design should be human-centered; the systematisms, hierarchy, and comprehensiveness in the process of index system establishment should be considered fully. The simulation experiment of complex interface design (CNC machine tool operating system) was carried out. Combined with the structural characteristics of complex interface design and ergonomics design criteria, the elements affecting the ergonomic reliability of working interface design were systematically collected by sending questionnaires to industrial design and mechanical design experts and students
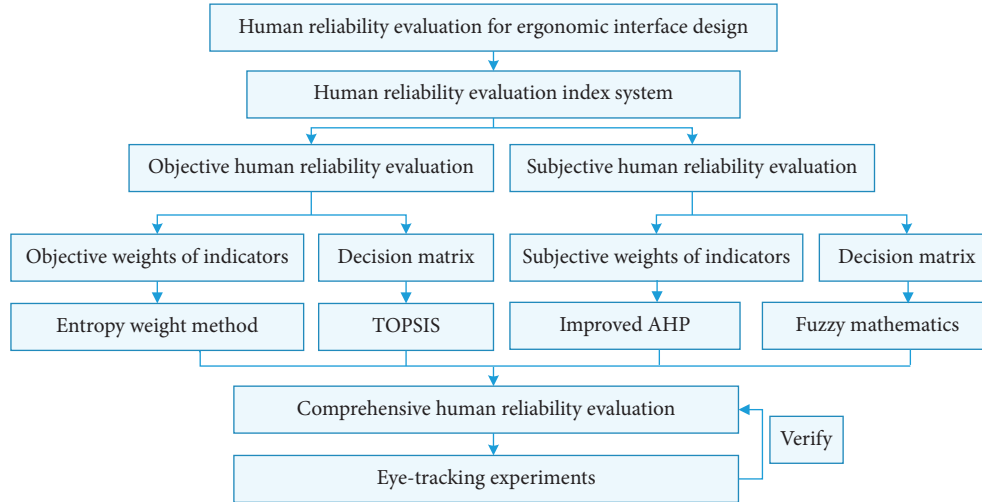
FIGURE 1: The process used for human reliability evaluation based on the objective and subjective method.

in addition to the CPCs provided by CREAM. But CREAM mainly focuses on cognitive error model and framework and the factor "Adequacy of MMI (Man-Machine Interface) and operational support" is just one condition of CPCs [31], so it is not very suitable for the systematic and comprehensive evaluation targeted ergonomic interface design. The index system of human reliability evaluation is improved based on CPCs to be suitable for the objective and subjective comprehensive method. There are altogether 26 network questionnaires sent out and obtained in about one week. Combined with some references, the human reliability evaluation indexes for complex interface design are listed in Table 1.

2.2. Objective Weights Determined by Improved Entropy-TOPSIS. The entropy weight method is a kind of objective weighting method, while the TOPSIS evaluation method has an objective and fair feature. Those two methods can be complementary to the expert-dependent and subjective AHP. The essence of entropy is the expected value of information. According to the calculation characteristic of Theil's entropy, it is especially suitable for solving multilevel weight calculation problems. TOPSIS was first proposed by C. L. Wang and K. YOON in 1981. TOPSIS is a kind of ranking method according to the proximity of assessment objects with a limited number to the ideal target, which is the assessment method of relative merits among the existing objects [32]. TOPSIS is ranked according to the distance detected from the best solution and the worst solution to the assessment object, respectively. The result is optimal, if the assessment object is the furthest from the worst solution while it is the closest to the best solution. On the contrary, it is not optimal. In addition, every index value of the best solution reaches the best value of every assessment index, while every index value of the worst solution reaches the worst value of every assessment index. TOPSIS can make sufficient analysis based on the information from the raw data, so the gaps between the different assessment programs can be reflected accurately from the results analyzed by

TOPSIS. Through the combination of Theil's entropy and TOPSIS, objective evaluation can be performed well.

The core idea of the objective weight determined by the entropy weighting method is based on the variability of indexes [33]. Suppose U is the probability of event A going to happen, $P(A) = U$. The amount of information that happens in this event is $E(U)$, which must be a minus function of U. So, the formula is given as $E(U) = \log(1/U)$. When there are n possible events $1, 2, \ldots, n$, the corresponding probability hypotheses are $U_1, U_2, \ldots, U_n$, respectively, and $\sum_{i=1}^{n} U_i = 1$.

Entropy or expected information can be regarded as the sum of the product of every piece of information in addition to their corresponding probabilities:

$$E(U) = \sum_{i=1}^{n} U_i h(U_i) = \sum_{i=1}^{n} U_i \log\left(\frac{1}{U_i}\right). \tag{1}$$

In order to measure the contribution of the data in the same group and with the other group to the total gap, Theil's entropy is used [34]:

$$T = \log n - E(U). \tag{2}$$

Since $\sum_{i=1}^{n} U_i = 1$,

$$\log n = \sum_{i=1}^{n} U_i \log n. \tag{3}$$

Substitute formula (3) and formula (1) into formula (2); then,

$$T = \log n - E(U) = \sum_{i=1}^{n} U_i \log n - \sum_{i=1}^{n} U_i \log\left(\frac{1}{U_i}\right)$$

$$= \sum_{i=1}^{n} U_i \left[\log n - \log\left(\frac{1}{U_i}\right)\right] = \sum_{i=1}^{n} U_i (\log n U_i). \tag{4}$$

a_i represents the number of mistakes an expert makes during operating the interface according to the i-th index, then U_i is the share of the error number based on the i-th index in the

TABLE 1: The index system of human reliability evaluation for complex interface design.

Target layer	Criterion layer	Index layer
The index system of human reliability evaluation	C_1 misperception	L_{11} fuzzy semantics; L_{12} limited vision; L_{13} optical illusion; L_{14} weak visibility; L_{15} cognitive biases
	C_2 memory lapse	L_{21} thinking load; L_{22} forgotten; L_{23} false memories; L_{24} inaccurate recall; L_{25} lack of memory assistance; L_{26} time pressure
	C_3 carelessness	L_{31} diminished intention; L_{32} overconfidence; L_{33} interference; L_{34} nervous at a loss; L_{35} information overload
	C_4 improper decisions making	L_{41} wrong planning target; L_{42} information misunderstanding; L_{43} the empirical; L_{44} no plan; L_{45} plan missing; L_{46} reasonably planned but not executed according to the plan
	C_5 improper operation	L_{51} inadequate training; L_{52} improper operating procedure; L_{53} taking shortcuts; L_{54} habitual control; L_{55} illegal manipulation; L_{56} lack of operational knowledge; L_{57} wrong operation time
	C_6 improper interface layout design	L_{61} improper functional partitioning; L_{62} indistinguishable different information; L_{63} poor logic of information arrangement; L_{64} unreasonable interface hierarchy; L_{65} unreasonable interface link order
	C_7 improper icon and text display design	L_{71} poor readability of characters; L_{72} inscrutable icon; L_{73} inscrutable jargon; L_{74} the icon not matching the task properly; L_{75} jargon not matching the task properly; L_{76} contradictory tag
	C_8 improper prompt feedback	L_{81} no prompt information; L_{82} unreasonable time of feedback response; L_{83} indistinguishable feedback; L_{84} perceived redundant information; L_{85} wrong feedback prompt

total number of errors $U_i = a_i / \sum_{i=1}^{n} a_i$, and the average value $\bar{a} = \sum_{i=1}^{n} a_i / n$. So, formula (4) can be expressed as

$$T = \sum_{i=1}^{n} \frac{a_i}{\sum_{i=1}^{n} a_i} \left(\log n \frac{a_i}{\sum_{i=1}^{n} a_i} \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{a_i}{\sum_{i=1}^{n} a_i / n} \left(\log \frac{a_i}{\sum_{i=1}^{n} a_i / n} \right) = \frac{1}{n} \sum_{i=1}^{n} \frac{a_i}{\bar{a}} \log \frac{a_i}{\bar{a}}. \quad (5)$$

The feature weight of the i-th index according to the j-th participant is given by

$$p_{ij} = \frac{a_{ij}}{\sum_{i=1}^{n} a_{ij}}, \quad j = 1, 2, \dots, m. \quad (6)$$

The i-th index in terms of Theil's entropy is shown as

$$e_i = \frac{1}{n} \sum_{i=1}^{n} \frac{a_{ij}}{p_{ij}} \log \left(\frac{a_{ij}}{p_{ij}} \right). \quad (7)$$

This kind of generalized entropy can make the explanatory power of the figures in the other group and in the same group clearer on the total gap. Based on Theil's entropy, the weight of every index is

$$w_i^2 = \frac{1 - e_i / \sum_{i=1}^{n} e_i}{\sum_{i=1}^{n} \left(1 - e_i / \sum_{i=1}^{n} e_i \right)}, \quad (8)$$

where m is the participant number and n is the index number. The variation degree of the index value is smaller which indicates that users are liable to make similar mistakes during interface operation, and the weight of a certain index based on Theil's entropy is bigger.

Based on this human reliability valuation index system, the fault tree was established, and the critical importance degree was taken to build the evaluation decision matrix, which was used to be analyzed by TOPSIS. Just by the occurrence probability of the top event in the fault tree, the fact cannot be reflected that it is easier to reduce the occurrence probability of a highly probable base event than to decrease the occurrence probability of a basic event with low probability, so the critical importance degree is also analyzed deeply. In order to analyze how changes in the probability of occurrence of fundamental events will affect the probability of occurrence of top events, the probability importance of fundamental events should be calculated $T_f(i)$. The probability function f of the probability of the top event occurring is a multiple linear function. By taking a partial derivative of the independent variable p_i, the probability importance coefficient of the basic event is expressed as follows [35]:

$$T_f = \frac{\partial f}{\partial p_i}. \quad (9)$$

The importance degree of the basic event is indicated by the proportion between the relative rate of change about the probability of the basic event occurring and that of change about the probability of the top event occurring, which means the importance criteria of every basic event is based on the probability and the sensitivity of the event occurrence about itself. So, the critical importance is represented as [30]

$$T_F(i) = \frac{\partial \ln f}{\partial \ln p_i} = \frac{\partial f / f}{\partial p_i / p_i}. \quad (10)$$

The relation between probability important degree and critical importance is shown as follows:

$$T_F(i) = \frac{p_i}{f} T_f(i). \quad (11)$$

where $T_f(i)$ is probability important degree; p_i is the occurrence probability of an event; and f is the overall event

occurrence probability. The objective comprehensive assessment not normalized is represented as follows:

$$d_{ki} = 1 - \frac{T_F(i)}{\sum_{i=1}^{N} T_F(i)}, \quad (12)$$

where N represents the total number of all indicators of the base event in the fault tree.

There are k interfaces to be evaluated, and the standardized matrix of n evaluation indicators in every intermediate event of the fault tree is as follows:

$$\mathbf{D} = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{k1} & d_{k2} & \cdots & d_{kn} \end{bmatrix}. \quad (13)$$

The human reliability evaluation decision matrix built by the critical importance degree is analyzed by TOPSIS as follows.

The maximum value is defined as follows:

$$\begin{aligned} D^+ &= (D_1^+, D_2^+, \dots, D_i^+) \\ &= (\max\{d_{11}, d_{21}, \dots, d_{k1}\}, \max\{d_{12}, d_{22}, \dots d_{k2}\}, \dots, \max\{d_{1n}, d_{2n}, \dots d_{kn}\}). \end{aligned} \quad (14)$$

The minimum value is defined as follows:

$$\begin{aligned} D^- &= (D_1^-, D_2^-, \dots, D_i^-) \\ &= (\min\{d_{11}, d_{21}, \dots, d_{k1}\}, \min\{d_{12}, d_{22}, \dots d_{k2}\}, \dots, \min\{d_{1n}, d_{2n}, \dots d_{kn}\}). \end{aligned} \quad (15)$$

The distance between the $k-$th $(k = 1, 2, \dots, l)$ assessment indicator value and the maximum value is defined:

$$L_k^+ = \sqrt{\sum_{i=1}^{n} w_{ij} (D_i^+ - d_{ki})^2}. \quad (16)$$

The distance between the $k-$th $(k = 1, 2, \dots, l)$ assessment indicator value and the minimum value is defined:

$$L_k^- = \sqrt{\sum_{i=1}^{n} w_{ij} (D_i^- - d_{ki})^2}. \quad (17)$$

Then, the closeness of the ideal solution about the $k-$th $(k = 1, 2, \dots, l)$ evaluation indicator is shown:

$$S_k = \frac{w_i L_k^-}{w_i L_k^+ + w_i L_k^-}. \quad (18)$$

The objective human reliability evaluation vector can be calculated:

$$F_k = \frac{(1 - T_{fk}) \times S_k}{\sum_k^l (1 - T_{fk}) \times S_k}. \quad (19)$$

2.3. Subjective Weights Determined by Improved AHP.

AHP is often applied to complex unstructured decision-making problems; it is very suitable to deal with the problem of multiple elements, multiple criteria, or multiple levels. Human judgment is inherently ambiguous to some extent, so the method of fuzzy mathematics can be applied in building the uncertainty of people's perceptual thinking mode. Since the assessment figures are from the rational data collection and perceptual expert analysis, AHP improved by expert rank based on Kendall's coordination coefficient is selected to make the subjective evaluation. The traditional AHP has certain limitations, the evaluation indexes should not be too many; otherwise, the difference between the judgment matrix and the consistent matrix will be greatly based on the traditional AHP. But actually, there are too many comparative evaluation indexes in the human reliability evaluation of ergonomic interface design, and those comparative evaluation indexes may lead experts to be confused very easily based on the traditional AHP-Fuzzy method. Own to the too long time spent on comparative, the application process is not organized clearly enough. In order to determine the weight coefficient, the importance scale method proposed by Saaty is applied in the traditional AHP generally, which means that the evaluation index pair comparison method is adopted to assign an index of the same system to another index with an integer multiple of importance; this cannot show the fuzziness of human judgments owing to the integer assignment comparing two pairs of evaluation indexes. In addition, when there are a large number of elements to be compared in the traditional AHP, the importance scale may exceed people's psychological endurance. The improved AHP can solve the problem. While making it easier for experts to make decisions, the computation of AHP can be reduced. In addition, a part of computation about the judgment matrix testing and adjusting for consistency can be reduced, so the number of the times of consistency checks performed by experts can be reduced, so it can save the time of expert analysis. By

calculating Kendall's coordination coefficient, the consistency level of evaluation can be measured scientifically and objectively. In the whole calculation process of improved AHP-Fuzzy, only one step consistency test is needed to judge whether the ranking of experts tends to be the same.

Because expert ranking data is the analysis of multisample related data, Kendall coordination coefficient W is generally used to test the consistency of ranking analysis. According to the indicators given by m experts, each indicator importance in the h criterion layer $B_1^{(h)}, B_2^{(h)}, \ldots, B_k^{(h)}$ is arranged from small to large. The rank is $1, 2, \ldots, k$.

H_0:p group assessment is unrelated or random m, and H_1:p group assessment is positively correlated or more or less consistent.

Kendall coefficient of concordance [36]:

$$W = \frac{12S}{m^2(n^3 - n)}, \tag{20}$$

$$S = \sum_{i=1}^{n} \left(R_i - \frac{m(n+1)}{2} \right)^2, \tag{21}$$

where m represents judge number, n represents object number, and R_i is the sum of ranks of the i-th object. When there are e identical data, the corrected statistical W_c is shown as follows:

$$W_c = \frac{12 \sum_{i=1}^{n} R_i^2 - 3m^2 n(n+1)^2}{m^2(n^3 - n) - m \sum_{k=1}^{e} \left(t_k^3 - t_k \right)}, \tag{22}$$

where t_k is the number of tied ranks about each k in e groups of ties. If $n \leq 7$ and $m \leq 20$, the critical values table for W is recommended to use [37]. If the n and m values exceed the Kendall coordination coefficient W value table, the chi-square values are calculated by the large sample approximation method:

$$\chi^2 = m(n-1)W. \tag{23}$$

After getting the chi-square value, according to the freedom degree $v = n - 1$, consult the chi-square threshold table to check whether the ranking results of m experts are consistent. If the critical value is $\chi_{0.05, n-1}^2 < \chi^2$, then $p < 0.05$, rejecting the H_0 hypothesis and accepting H_1, which means the ranking of each index in the n criterion layers by m participants is consistent.

By using AHP, a complex problem can be simplified to an orderly hierarchical structure. The criterion layer includes $B_1^{(h)}, B_2^{(h)}, \ldots, B_k^{(h)}$ and other factors. When the number of evaluation indicators is large, it is very difficult and time-consuming to use traditional AHP to carry out multiple comparisons for evaluation indicators. Therefore, based on improving experts' quantitative analyses to obtain the average rank, this paper uses the AHP model to obtain the index weights in the criterion layer.

Based on the basic principle of AHP, the judgment matrix is constructed, and the lowest judgment matrix of each index in the same system is constructed as follows [38]:

$$B_k^{(h)} = \frac{\sum_{i=1}^{m} n_i}{m}, \tag{24}$$

$$B^{(h)} = \begin{pmatrix} 1 & \dfrac{B_2^{(h)}}{B_1^{(h)}} & \cdots & \dfrac{B_k^{(h)}}{B_1^{(h)}} \\ \dfrac{B_1^{(h)}}{B_2^{(h)}} & 1 & \cdots & \dfrac{B_k^{(h)}}{B_2^{(h)}} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{B_1^{(h)}}{B_k^{(h)}} & \dfrac{B_2^{(h)}}{B_k^{(h)}} & \cdots & 1 \end{pmatrix} = \left(\dfrac{B_n^{(h)}}{B_m^{(h)}} \right)_{k \times k}, \tag{25}$$

$$\left(m, n = 1, 2, \ldots, k; B_k^{(h)} = 1, 2, \ldots, k \right).$$

In the previous formula, $B_k^{(h)}$ represents the average rank of participate ranking. The basic judgment matrix (25) has the following properties:

$(B_n^{(h)}/B_n^{(h)}) = 1, (B_n^{(h)}/B_m^{(h)}) = 1/(B_m^{(h)}/B_n^{(h)}) > 0$. Because $B_n^{(h)}/B_k^{(h)} \times B_k^{(h)}/B_m^{(h)} = B_n^{(h)}/B_m^{(h)}$, the matrix conforms to the condition of consistency matrix; that is, no more consistency checking is needed.

Since $B^{(h)}$ is a consistency matrix, the rank of $B^{(h)}$ is 1, the unique nonzero eigenvalue of $B^{(h)}$ is k, any column vector of $B^{(h)}$ is the eigenvector corresponding to k, and the normalized eigenvector of $B^{(h)}$ can be used as the weight vector of $B^{(h)}$; that is,

$$B^{(h)}w = kw, w = (w_1, w_2, \ldots w_k). \tag{26}$$

After the w normalization, the obtained vector $w^{(h)}$ is the relative importance of every element in the criterion layer of the lowest layer:

$$W_j^{(h)} = \begin{cases} \dfrac{1 - w_j}{\sum_{j=1}^{k} \left(1 - w_j \right)} & w_j \geq 0, \\[4mm] \dfrac{1 + w_j}{\sum_{j=1}^{k} \left(1 + w_j \right)} & w_j < 0. \end{cases} \tag{27}$$

After calculating the relative importance of the elements at each layer, start at the top layer (the overall objective), and the comprehensive importance of the elements at each layer can be obtained from the top to the bottom. This calculation needs to be carried out layer by layer from top to bottom in order.

Assuming that we are now at the h layer, a subsystem contains k elements $B_1^{(h)}, B_2^{(h)}, \ldots, B_k^{(h)}$; the weight coefficients of each element are $w_1^{(h)}, w_2^{(h)}, \ldots, w_k^{(h)}$; and then in the next layer of $B_j^{(h+1)}$ corresponding to $w_j^{(h)}$ the relative importance of $B_{j1}^{(h+1)}, B_{j2}^{(h+1)}, \ldots, B_{jn}^{(h+1)}$ is $w_{j1}^{(h+1)}, w_{j2}^{(h+1)}, \ldots, w_{jn}^{(h+1)}$. If $B_j^{(h+1)}$ is not related to $B_k^{(h)}$, then $w_j^{(h+1)} = 0$, and $w_j^{(h)} = \sum_{i=1}^{n} W_{ji}^{(h+1)}$ $(j = 1, 2, \ldots, k)$.

Therefore, the weight coefficient $W_{ji}^{(h+1)}$ of element $B_{ji}^{(h+1)}$ is

$$W_{ji}^{(h+1)} = w_{ji}^{(h+1)} w_j^{(h)}. \tag{28}$$

Because the basic layer matrix itself conforms to the nature of the consistency matrix, the consistency index of hierarchical ranking is 0. According to the principle of AHP, the consistency index of hierarchical total ranking must also be 0, which can subtract the steps of consistency checking and adjusting and save the computer computation. In order to show the uncertainty of people's subjective feelings, fuzzy mathematics principle is introduced to analyze the data processed by improved AHP:

$$\mathbf{Z} = \mathbf{W}_j \mathbf{C}$$

$$= (w_1, \ldots, w_j) \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1v} \\ c_{21} & c_{22} & \cdots & c_{2v} \\ \vdots & \vdots & \cdot & \vdots \\ c_{j1} & c_{j2} & \cdots & c_{jv} \end{bmatrix}, \tag{29}$$

$$c_{jv} = \frac{g_{jv}}{q}, \tag{30}$$

C is a fuzzy evaluation matrix, $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_l)$ represents the fuzzy comprehensive evaluation result, $Z_h (h = 1, 2, \ldots, l)$ reflects the position of the h-th decision in the overall decisions; g_{jv} is the number of all the votes for the v-th level of the j-th element about the interface operation evaluation; and q is the number of subjects joined in voting.

The result of every combined factor is given as

$$\mathbf{F}_z = \mathbf{Z}\mathbf{V}^T, \tag{31}$$

where V is the evaluation grade level vector and Z is the qualitative evaluation matrix.

3. Application Example of Human Reliability Evaluation in Control Interface Design

3.1. Objective Evaluation of the Interface Design. To analyze the human reliability of 3 different ergonomic interface designs about the CNC milling machine, 35 volunteers who were well-rested in good spirits participated in the operation experiments of three high-fidelity simulation interface designs. Among them, more than 25 people have experience in operating actual milling machines, and 6 people have significantly more experience in operating than others. From the total number of operational steps per operator and the number of operational and attentional lapses per operator based on each evaluation indicator, according to formula (1) to formula (8), the results about objective weights of 3 interface designs are shown, respectively, in Figure 2. In Figure 2(a), the indicator weights in the criterion layer are shown and in Figure 2(b), the indicator weights in the index layer are shown. Based on the entropy weight calculation method, if multiple operators have similar behaviors or make similar mistakes for the same indicator, the weight value of this index will be greater. On the contrary, due to the different individual factors of the operator, the behavior is not similar, and the weight value will be smaller. In Figure 2(a), about design scheme A, the weight value of C_1 is the largest, which means operators are prone to make similar errors under the indicator about false perception. About design scheme B, the weight value of C_3 is the largest, which means operators are prone to make similar errors under the indicator about negligence. About design scheme B, the weight value of C_4 is the largest, which means operators are prone to make similar errors under the indicator about decision-making failure. About scheme A, the weight value of C_8 is the smallest. About scheme B, the weight value of C_8 is the smallest, too. About scheme C, the weight value of C_7 is the smallest.

According to formula (9) to formula (12), in order to make the decision matrix more scientific, the analysis method of the fault tree is introduced to process the original data. In Figure 3(a), it is the critical importance value of every indicator in the index layer after normalized and positive processes building the decision matrix. Then, the calculation method of TOPSIS is introduced, in Figure 3(b), and the differences between the indicator value and the optimal value about the same indicator in the index layer are represented, while the differences between the indicator value and the worst value about the same indicator in the index layer are represented in Figure 3(c). The difference values form the decision matrix of the objective evaluation.

According to formula (9), the occurrence probabilities of the top event which are also the probabilities of operation failure of three interface design schemes are $T_f = (0.1567, 0.1453, 0.0887)$.

3.2. Subjective Evaluation of the Interface Design. In human reliability analysis, the subjective factors cannot be ignored. There are A, B, C three types of CNC system simulation for control interface design. 26 participants were investigated, including 15 that have rich experience in machine tool use and 11 that are beginners. The importance ranking of the man-machine reliability evaluation index is carried out from light to heavy. Because it exceeds the query range of the Kendall W coordination coefficient value table, the chi-square test is adopted to check the consistency.

As in Table 2, Kendall's $W = 0.363, \chi^2 = 88.931$. According to the freedom degree $v = n - 1 = 8 - 1 = 7$, check the consistency of the ranking results of 26 experts by checking the chi-square boundary table. Critical value $\chi^2_{0.05, 7} = 14.067 < \chi^2$, then $p < 0.05$, the H_0 hypothesis is rejected, and H_1 is accepted.

After 4 rounds of coordination, 26 experts ranking all of the indicators in the index layer and criterion layer achieve consistency. In Figure 4, sequences of indicators in the index layer and criterion layer are shown. In the criterion layer, the 5th index, improper operation, is the most important according to the experts' ranking. The 2nd index, memory lapse, is the least important among all the indicators. Based on the indicator sequences, according to formula (24) to
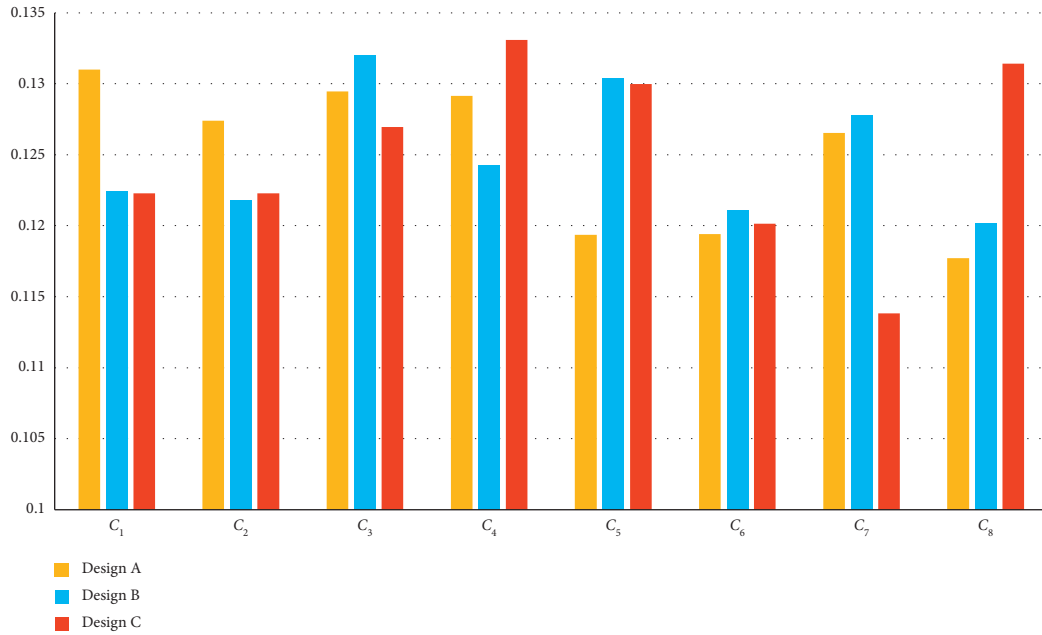
(a)
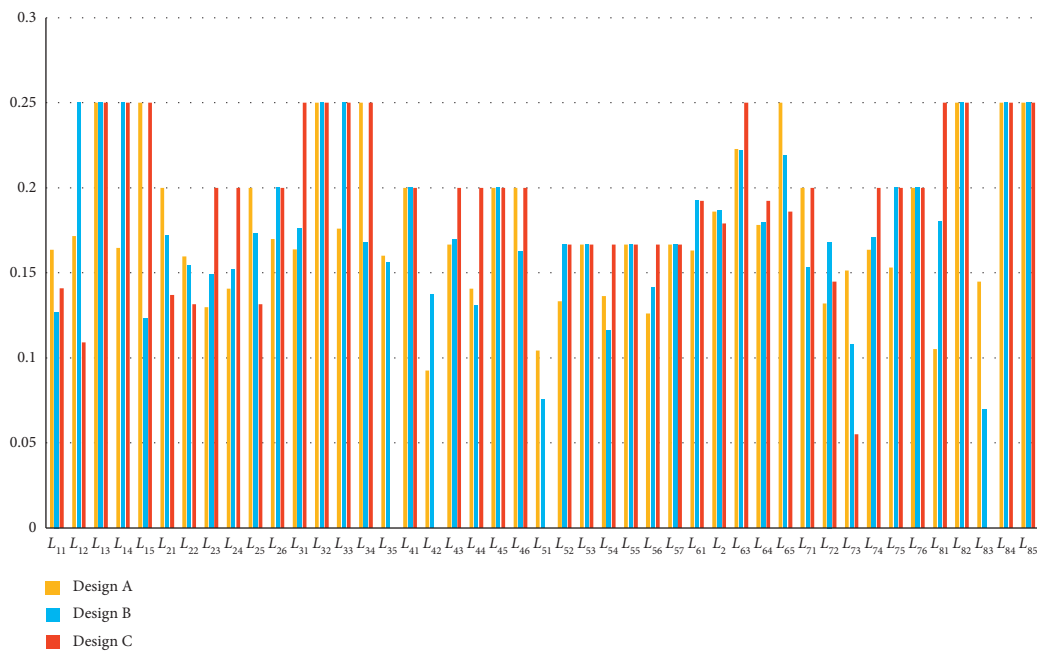
(b)

FIGURE 2: Objective weights calculated by entropy weight method. (a) Indicators in the criterion layer. (b) Indicators in the index layer.

formula (28), the subjective weights calculated by improved AHP are shown in Figure 5. The larger number of the indicator sequence is, the greater the weight is. So, the index weight of improper operation is the greatest, while the index weight of memory lapse is the smallest in the criterion layer. There are some differences from the index weights analyzed by the objective human reliability evaluation method owning to the objective evaluation results based on the results of objective experimental data collection. The objective weights of the same index based on the different interaction schemes operating are different because of the

objective weights affected by objective experimental data under different circumstances. Otherwise, the subjective weights of the same index are the same under different circumstances owning to the ranking of indexes by experts not considering the different interaction schemes. In this way, the subjective method and objective method can play a complementary role, and the human reliability evaluation based on subjective and objective methods is more systematic so that the final results are more reliable.

35 volunteers gave the number of invalid minds and behaviors based on every indicator under 5 grand levels
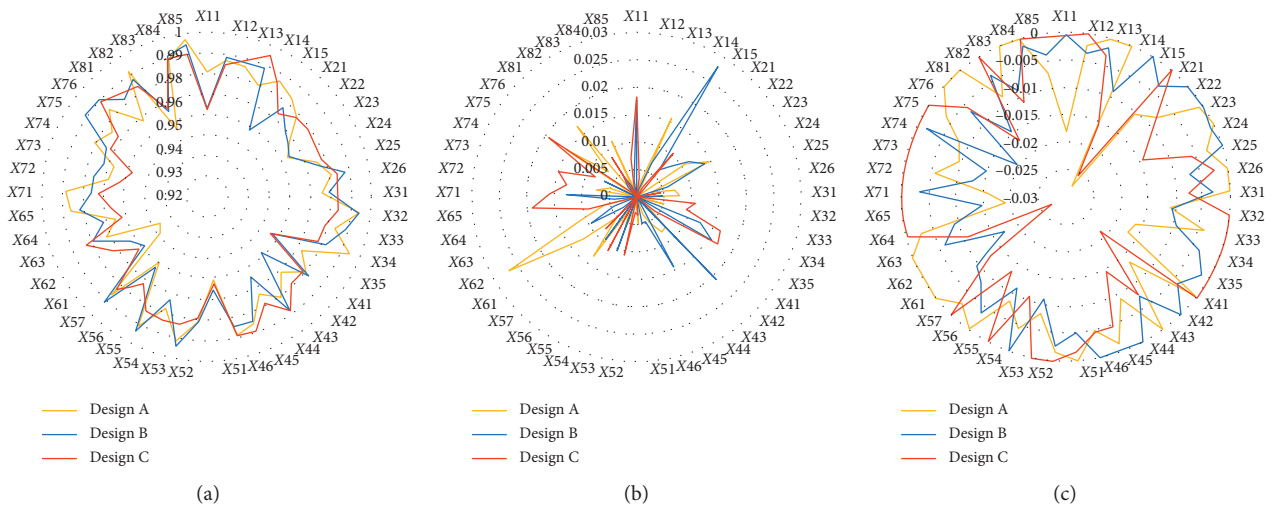
(a)

(b)

(c)

Figure 3: The critical degree values of importance processed and the best and the worst difference values in the index layer. (a) Critical degree values of importance processed. (b) Differences between the indicator values and the optimal values. (c) Differences between the indicator values and the worst values.

Table 2: Average rank of man-machine reliability to evaluation criterion layer $C^{(1)}$.

$C_1^{(1)}$	$C_2^{(1)}$	$C_3^{(1)}$	$C_4^{(1)}$	$C_5^{(1)}$	$C_6^{(1)}$	$C_7^{(1)}$	$C_8^{(1)}$
3.67	2.50	2.66	5.49	5.93	5.79	5.87	4.10

- -*- Values in misperception index layer
- -○- Values in memory lapse index layer
- ⋯□⋯ Values in carelessness index layer
- Values in improper decision-making index layer
- -■- Values in improper operation index layer
- Values in improper interface layout design index layer
- Values in improper graphic and text display design index layer
- Values in improper prompt feedback index layer
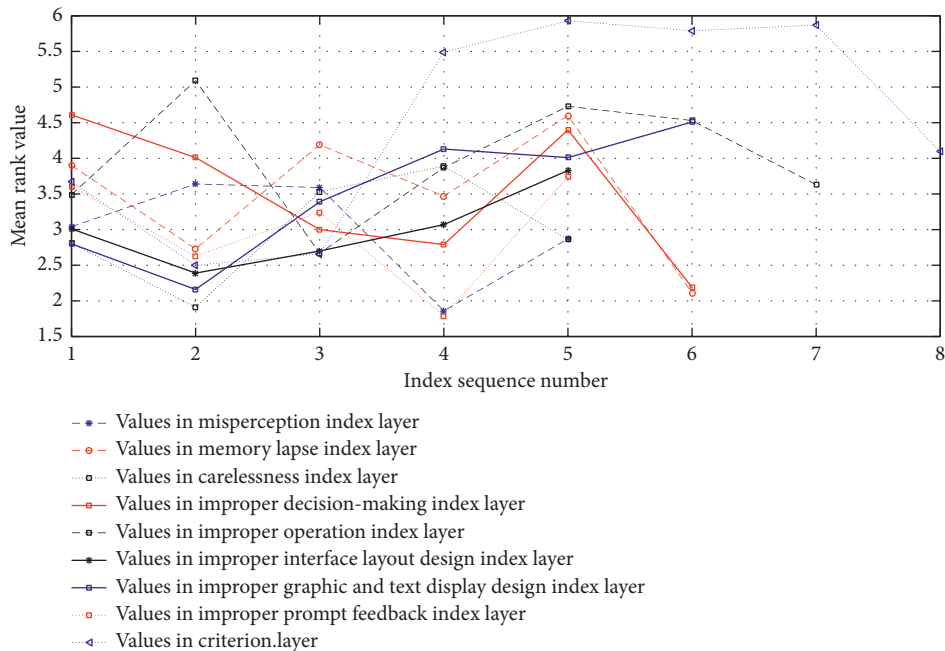- ⋯◁⋯ Values in criterion.layer

Figure 4: Sequences of indicators in the index layer and criterion layer.

according to their own perceptual knowledge after operation experiments. According to formula (29), in Table 3, the results of the subjective reliability assessment matrix reflecting 5 grand levels built are listed.

3.3. Subjective and Objective Comprehensive Evaluation Results. Owning to the objective weights calculated by Theil's entropy, the objective weights of the same indexes base on types A, B, and C are different. So, it is hard to
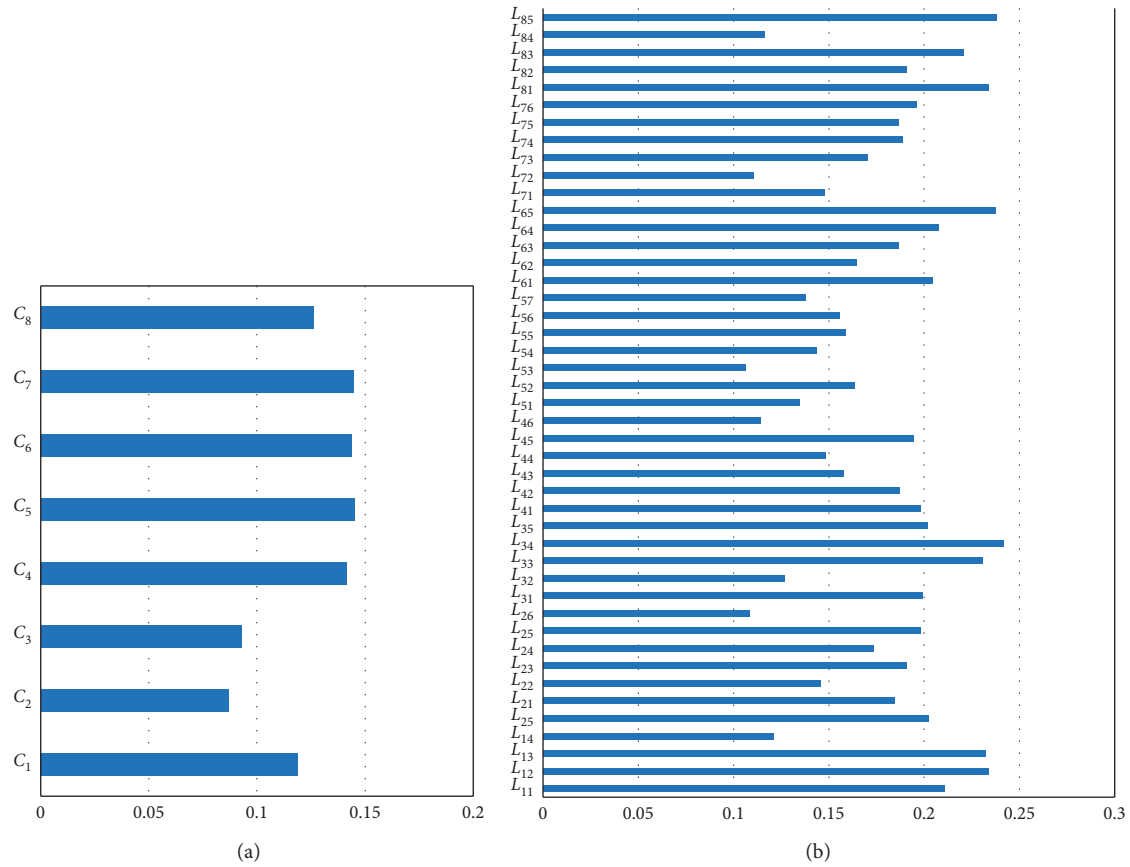
FIGURE 5: Subjective weights calculated based on improved AHP. (a) The criterion layer indicators. (b) The index layer indicators.

TABLE 3: Reliability evaluation matrix based on 5 grand levels given by fuzzy about types A, B, and C.

Interaction design	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
	0.006017	0.005445	0	0	0	0.005924	0.00486	0
	0.017575	0.034598	0	0.009586	0.012391	0.033267	0.012246	0.012989
A	0.052163	0.044445	0.024128	0.032266	0.032672	0.044944	0.032801	0.018444
	0.120065	0.127449	0.135788	0.142011	0.146856	0.136076	0.120256	0.097355
	0.804181	0.749951	0.840084	0.816137	0.808081	0.779789	0.829836	0.871213
	0.006017	0.004166	0	0.007754	0.012959	0.010534	0	0.006678
	0.029428	0.018748	0.021665	0.022904	0.029379	0.034496	0.027498	0.048238
B	0.019615	0.027789	0.045794	0.031393	0.054302	0.070879	0.046772	0.041192
	0.107365	0.202221	0.09891	0.165524	0.106956	0.159912	0.137176	0.107353
	0.837575	0.747076	0.833631	0.772426	0.796404	0.776849	0.788554	0.796538
	0	0	0	0	0	0	0	0
	0.0127	0.01509	0	0.005341	0.003843	0	0.00486	0
C	0.028571	0.028571	0.034338	0.028571	0.032414	0.023239	0.008016	0.00631
	0.070832	0.086284	0.10497	0.081464	0.099277	0.116021	0.116323	0.081897
	0.887897	0.870054	0.860693	0.884623	0.864466	0.86074	0.8708	0.911793

determine the best scheme just according to Figure 6(a) and Figure 6(b). But according to Figure 6(c) which shows the human reliability evaluation values in the criterion layer based on obtaining the subjective evaluation results of the index layers, the best scheme which is type C can be distinguished preliminarily. According to formula (16) and formula (17), Figures 6(a) and 6(b) are obtained, respectively. Based on formula (18), the indicator weights in the criterion layer, the vector of the degree close to the ideal

solutions is obtained as $S_k = (0.551631449, 0.501348394, 0.537021162)$. By formula (19), $F_k = (0.465203139, 0.428510755, 0.489400885)$. After normalization, the objective weight vector is obtained as $F_k = (0.336344565, 0.309815759, 0.353839676)$. From the objective data analysis results, the human reliability and safety evaluation of type C is the best one among the three interface design schemes. In addition, according to equation (31), the fuzzy assessment matrix of the subjective human reliability evaluation as
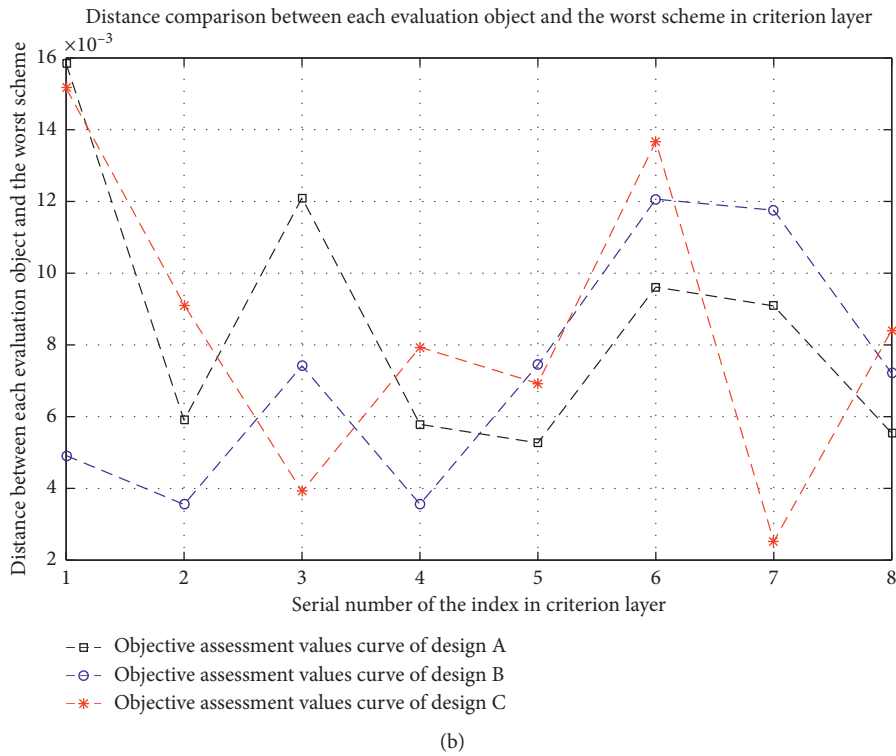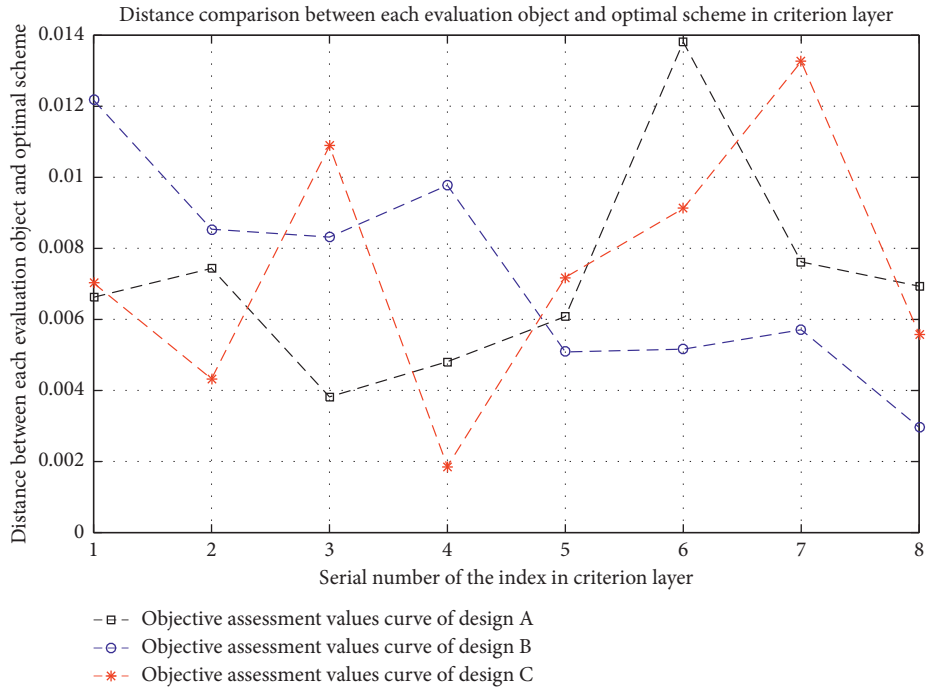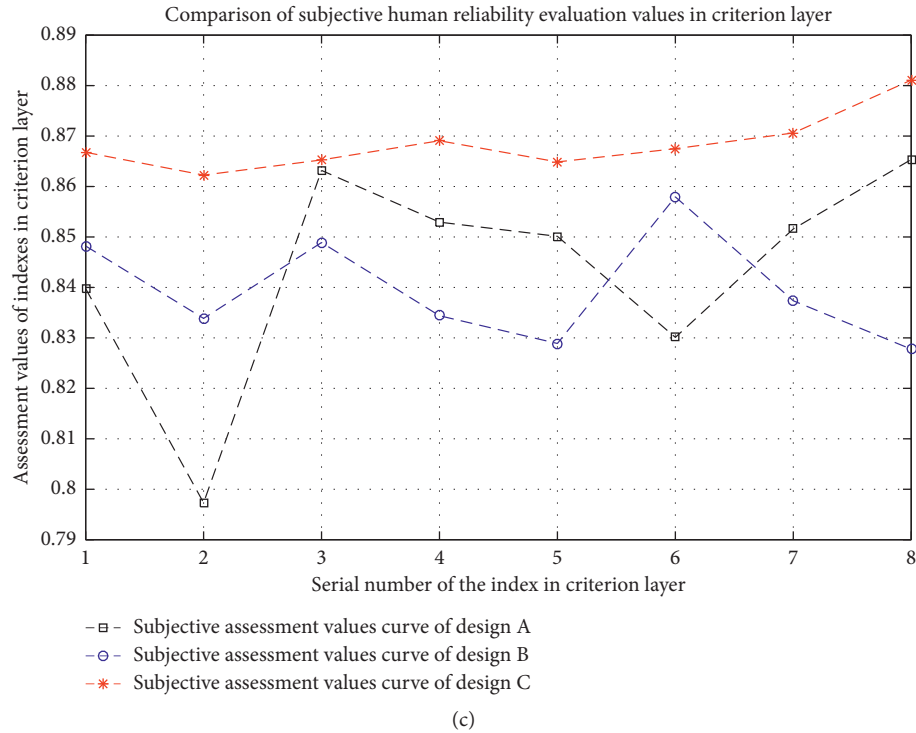
Distance comparison between each evaluation object and optimal scheme in criterion layer

- –□– Objective assessment values curve of design A
- –○– Objective assessment values curve of design B
- –*– Objective assessment values curve of design C

(a)

Distance comparison between each evaluation object and the worst scheme in criterion layer

- –□– Objective assessment values curve of design A
- –○– Objective assessment values curve of design B
- –*– Objective assessment values curve of design C

(b)

Figure 6: Continued.

(c)

FIGURE 6: Objective distances and subjective human reliability evaluation decision matrix in criterion layer. (a) Distance between every evaluation object and optimal scheme. (b) Distance between every evaluation object and the worst scheme. (c) Comparison of subjective human reliability evaluation values.

shown in Figure 6(c) can be obtained according to the fuzzy matrix of Table 3 multiplied by evaluation grade level vector *V*. The grade vector is denoted as $V = (0.1, 0.3, 0.5, 0.7, 0.9)$. The subjective feeling of user operation interface C is obviously better than interfaces A and B. According to the indicator weights in the criterion layer, the total weight vector of the 3 types is $F_z = (0.84517, 0.83953, 0.86876)$. After normalization, $F_z = (0.33099, 0.328781, 0.340229)$, which means the human reliability and safety evaluation of type C is the best among the three types, and that of type A is better than type B.

4. Verification Based on Eye-Tracking Experiments

Eye movement data obtained by eye-tracking experiment can objectively explore the relationship between eye movement and people's psychological activities. Eye movements can reflect the selection patterns of visual information. It has great significance to reveal the psychological mechanism in the process of cognition. In the same situation, the subjects' choice orientation can be detected by recording the eye movement information, so the eye-tracking experiment can be used as a kind of typical ergonomic objective research method for human reliability analysis [39] and can be used to verify the validity and rational of results analyzed by an objective and subjective comprehensive method based on Entropy-TOPSIS-AHP.

There are 16 volunteers who took part in the experiment. Figure 7 shows three screenshots of the heat map when participants observe and controlled the interfaces of types A, B, and C, respectively. The heat map of attention can show how much subjects pay attention to the information presented on the interface and whether the information can attract the eyes of the subjects. The visual attention level is shown in Figure 7. The gaze trajectory, gaze time, scan time, scan path length in eye movement data were analyzed, and visual cognitive experiments of different error factors were carried out in different subinterfaces and subtask environments step by step. During the operation, the eyes of participants are not looking for target places, and the number of times about target not being observed and not found is counted as the number of attention failures. In this process, participants may be groping and unable to find the target operation object, or they may let their minds wander. The variance analysis is used based on the number of attention failures. The variance analysis results of attention failures in three different kinds of interface design are given. The size of the F statistic is 30.0500. By checking the critical value table of F distribution, $p \leq 0.001$. So, the null hypothesis is rejected that the mean values of types A, B, and C are all equal, whereas the alternative hypothesis is accepted that the mean values of types A, B, and C are not all equal. Those indicate that the means of P_x from types A, B, and C, respectively have a statistically significant difference. Meanwhile, the homogeneity of variance was tested by Bartlett test, then $\chi^2 = 4.0405$, and the corresponding $p = 0.133$, which means
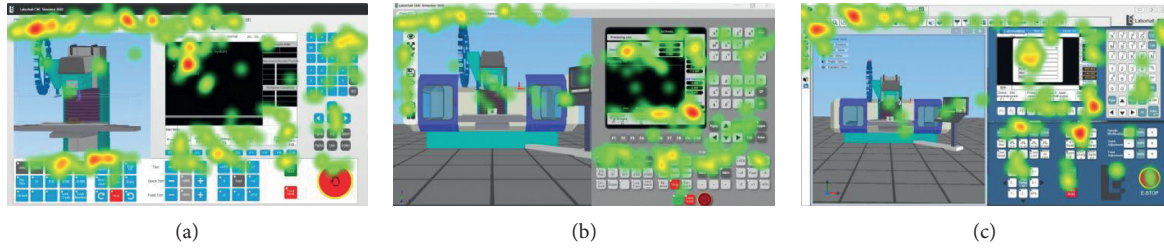
(a)												(b)												(c)

Figure 7: Heat map for the three interface design schemes of the simulation model. (a) Design A. (b) Design B. (c) Design C.

Table 4: Multiple comparisons of fixation point failure for three different types of interface design.

Interfaces I and II	$\overline{P_I} - \overline{P_{II}}$	p
B and A	7.0000	0.0010
A and C	6.3750	0.0020
B and C	13.3750	≤0.001

Table 5: Multiple comparisons between three kinds of interface design of finding cutting tool.

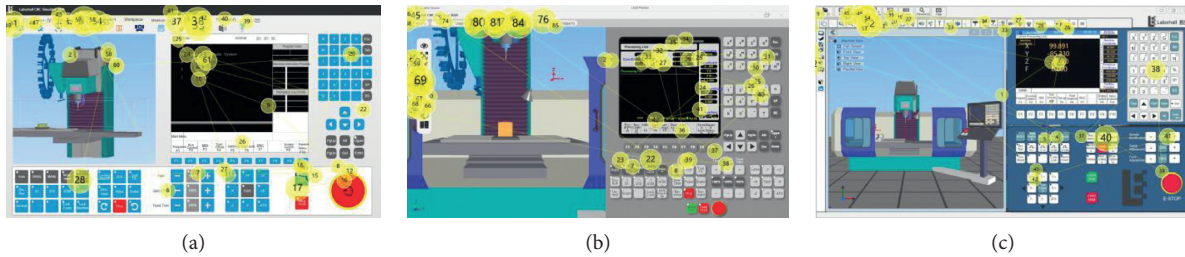| Interfaces I and II	$\overline{P_I} - \overline{P_{II}}$	p
B and A	4.4375	0.6480
A and C	12.2500	0.0040
B and C	16.6875	≤ 0.001

(a)												(b)												(c)

Figure 8: Gaze plot for three interface design schemes of the simulation model. (a) Design A. (b) Design B. (c) Design C.

that null hypothesis that the variance of all types is equal can be accepted under the condition of 15% test level. So, it is right to apply those data obtained from the experiments in the variance analysis.

To analyze the smallest mean of coefficient of variation and the largest mean of coefficient of variation among the three types, multiple comparisons are performed based on the Bonferroni method. The analysis of the multiple comparisons between the three types is shown in Table 4.

According to the multiple comparisons in Table 4, the mean of attention failures about design scheme B was 7.0000 higher than that of type A, and the P value was 0.0010, which means the result has statistical significance. The mean of attention failures of type C was 6.3750 lower than that of type A in addition to 0.0020 P value; the result has statistical significance too. The mean of attention failures of type C was 13.3750 lower than type B, $p \leq 0.001$, which means that the

result has statistical significance under the condition of 0.5% test level.

According to the observation sequence of experiment participants finding the cutting tool as shown in Figure 8, those are eye-tracking pictures based on just one of the subjects owning to the limited exhibition space. Those were chosen to be typical. The sequence numbers of all the subjects are analyzed by the variance analysis. The results are shown in Table 5.

The sequence number of finding the cutting tool based on design C is obviously smaller than design A and design B. But the sequence number of design A is not significantly different from that of design B. The analysis results of multiple comparisons indicate that users have more difficulty in concentrating when they use types A and B than when they use type C. The participants and experts can find the proper bottoms to control the interfaces used by type C

more effectively and efficiently than types A and B. In addition, the users can be more focused on the control interface when they use type C than when they use types A and B in a limited time. Some tools of types A and B are replaced by icons and those tools are hidden in the drop-down menu of the icon, but those icons are not commented, which makes it take a long time for users to find the target tools. In addition, those icons of type B have no popup instructions, and each operation module is too close to each other with no obvious distance, which leads to more attention lapses.

5. Conclusions

The objective and subjective comprehensive method of human reliability evaluation contributes to the stochastic behavior under operations. The working interface design always contains a large amount of information display, which requires users to process large-scale information in a limited time. Sometimes, users have a high cognitive difficulty in the process of execution, and it is easy to forget and misread. At the same time, the information content in interface design belongs to professional high-level cognitive information content. Users will have certain cognitive impairment when they perceive and understand, and users are more likely to make mistakes in the process of operation, which increases the difficulty of users' cognition and the error rate. It has important significance to build a human reliability evaluation model for complex working interface design considering the objective and subjective factors in the human-machine interaction process:

(1) According to the research results of the simulation experiments, questionnaire surveys, and references, the human reliability index system suitable for ergonomic interface design has been established. Human reliability analysis is affected by objective and subjective factors. On this basis, a more systematic and comprehensive human factor reliability evaluation method is proposed. The objective evaluation and the subjective evaluation are combined to reflect the fuzziness of human consciousness and strengthen the rational data analysis in human reliability analyses.

(2) Owning to the feature of human reliability evaluation with perception and reason, a comprehensive weighting method combining Entropy-TOPSIS-AHP is proposed to make the objective evaluation complementary to the subjective one. In addition, the objectivity and accuracy of the human reliability evaluation indexes at all levels are improved. The innovation of this search is that Theil's entropy is used for the objective evaluation owning to the easier calculation in the multilevel and multifactor structure. In addition, the improved AHP is used for the subjective evaluation owning to the greater efficiency and better execution in the importance order process for the evaluation factors than the expert judgment process of two factors contrast in the traditional AHP. The improved AHP solves the problem that there may be a big gap between the judgment matrix and consistency matrix when there are too many subjective evaluation indexes.

(3) Eye movement experiment as an objective ergonomic research method is used in recording human observation and human error data. The feasibility and effectiveness of the human reliability evaluation method based on the objective and subjective comprehensive method are verified by eye movement experiments. According to the evaluation results, feedback can be given to the control interface design to enhance ergonomic reliability and safety.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] S. Zhang, *Research on Mobile Terminal Interface Design of Special Population Based on Human Factor Reliability Analysis*, Xi 'an University of Architecture and Technology, Xi'an, China, 2013.

[2] T. J. Albin, "Measuring the validity and reliability of ergonomic checklists," *Work (Reading, Mass.)*, vol. 43, no. 3, p. 381, 2012.

[3] S.-P. Zhu, H.-Z. Huang, W. Peng, H.-K. Wang, and S. Mahadevan, "Probabilistic physics of failure-based framework for fatigue life prediction of aircraft gas turbine discs under uncertainty," *Reliability Engineering & System Safety*, vol. 146, pp. 1–12, 2016.

[4] S.-P. Zhu, Q. Liu, Q. Lei, and Q. Wang, "Probabilistic fatigue life prediction and reliability assessment of a high pressure turbine disc considering load variations," *International Journal of Damage Mechanics*, vol. 27, no. 10, pp. 1569–1588, 2018.

[5] Z. Liu, X. Liu, K. Wang et al., "GA-BP neural network-based strain prediction in full-scale static testing of wind turbine blades," *Energies*, vol. 12, no. 6, pp. 1–15, 2019.

[6] J. M. Rohani, A. J. Adeyemi, R. Abd Aziz et al., "The inter-rater and intra-rater reliability analysis of workplace ergonomic risk assessment," *Journal Technology*, vol. 80, no. 1, pp. 53–59, 2018.

[7] J.-K. Park, J. Boyer, J. Tessler et al., "Inter-rater reliability of PATH observations for assessment of ergonomic risk factors in hospital work," *Ergonomics*, vol. 52, no. 7, pp. 820–829, 2009.

[8] K. Norman, H. Alm, E. W. Tornqvist, and A. Toomingas, "Reliability of a questionnaire and an ergonomic checklist for assessing working conditions and health at call centres," *International Journal of Occupational Safety and Ergonomics*, vol. 12, no. 1, pp. 53–68, 2006.

[9] L. F. de Campos, J. N. de Lanutti, and L. C. Paschoarelli, "Product functions: interfaces with ergonomic design," *Work (Reading, Mass.)*, vol. 41, pp. 960–963, 2012.

[10] T. Mulyana, R. Ibrahim, and E. Abd Rahim, "Human machine interface design analysis of defect detection prototype by wonderware InTouch software," *Journal of Physics. Conference Series.*vol. 1150, no. 1, p. 12034, 2019.

[11] J. Jiang, L. Zhang, T. Xie et al., "Optimal design methods for a digital human-computer interface based on human reliability in a nuclear power plant Part 2: the optimization design method for component quantity," *Annals of Nuclear Energy*, vol. 106, pp. 247–255, 2017.

[12] B. Tian, S. Yu, J. Chu, and W. Li, "Shoulder girdle muscle activity and fatigue during automobile chassis repair," *International Journal of Occupational Medicine and Environmental Health*, vol. 32, no. 4, pp. 537–552, 2019.

[13] X. Liu, Z. Liu, Z. Liang, S.-P. Zhu, J. A. F. O. Correia, and A. M. P. De Jesus, "PSO-BP neural network-based strain prediction of wind turbine blades," *Materials*, vol. 12, no. 12, p. 1889, 2019.

[14] A. Ramezani, T. Nazari, A. Rabiee, K. Hadad, and M. Faridafshin, "Human error probability quantification for NPP post-accident analysis using Cognitive-Based THERP method," *Progress in Nuclear Energy*, vol. 123, p. 103281, 2020.

[15] M. Mirzaei Aliabadi, I. Mohammadfam, and K. Salimi, "Identification and evaluation of maintenance error in catalyst replacement using HEART technique under fuzzy environment," *International Journal of Occupational Safety and Ergonomics*, pp. 1–13, 2021.

[16] S. I. Ahn and R. E. Kurt, "Application of a CREAM based framework to assess human reliability in emergency response to engine room fires on ships," *Ocean Engineering*, vol. 216, p. 108078, 2020.

[17] M. Mahdy, "Weighted entropy measure: a new measure of information with its properties in reliability theory and stochastic orders," *Journal of Statistical Theory and Applications*, vol. 17, no. 4, pp. 703–718, 2018.

[18] S. Sandoval, A. Torres, and N. Obregon, "Tools for the implementation of proactive maintenance in urban sewer systems by the use of flooding reliability and entropy of information concepts," *Revista facultad de ingenieria-universidad de antioquia*, no. 65, pp. 152–166, 2012.

[19] S. B. El-Ladan and O. Turan, "Human reliability analysis-Taxonomy and praxes of human entropy boundary conditions for marine and offshore applications," *Reliability Engineering & System Safety*, vol. 98, no. 1, pp. 43–54, 2012.

[20] M. A. Mohammed, K. H. Abdulkareem, A. S. Al-Waisy et al., "Benchmarking methodology for selection of optimal COVID-19 diagnostic model based on Entropy and TOPSIS methods," *IEEE Access*, vol. 8, pp. 99115–99131, 2020.

[21] R. Nagpal, D. Mehrotra, and P. K. Bhatia, "Usability evaluation of website using combined weighted method: fuzzy AHP and entropy approach," *International Journal of System Assurance Engineering and Management*, vol. 7, no. 4, pp. 408–417, 2016.

[22] A. Petruni, E. Giagloglou, E. Douglas, J. Geng, M. C. Leva, and M. Demichela, "Applying Analytic Hierarchy Process (AHP) to choose a human factors technique: choosing the suitable Human Reliability Analysis technique for the automotive industry," *Safety Science*, vol. 119, pp. 229–239, 2019.

[23] P. Dehghanian, M. Fotuhi-Firuzabad, S. Bagheri-Shouraki, and A. A. Razi Kazemi, "Critical component identification in reliability centered asset management of power distribution systems via Fuzzy AHP," *IEEE Systems Journal*, vol. 6, no. 4, pp. 593–602, 2012.

[24] B. Jasra and S. K. Dubey, "Reliability assessment of component-based software system using fuzzy-AHP," *Advances in Intelligent Systems and Computing*, Springer Singapore, vol. 731, , pp. 663–670, 2018.

[25] Z. Liu, X. Liu, S.-P. Zhu et al., "Reliability assessment of measurement accuracy for FBG sensors used in structural tests of the wind turbine blades based on strain transfer laws," *Engineering Failure Analysis*, vol. 112, p. 104506, 2020.

[26] B. Imşek and F. Pakdil, "Building a graphical user interface for concrete production processes: a combined application of statistical process control and design of experiment," *Arabian Journal for Science and Engineering*, vol. 44, no. 5, pp. 4373–4393, 2019.

[27] X. Wu, C. Xue, and H. Wang, "E-C mapping model based on human computer interaction interface of complex system," *Journal of Mechanical Engineering*, vol. 50, no. 12, pp. 206–212, 2014.

[28] Z. Liu, Y. Le, Y. Li et al., "Comparisons of two non-probabilistic structural reliability analysis methods for aero-engine turbine disk," *International Journal of Turbo & Jet-Engines*, vol. 3, no. 34, pp. 295–303, 2017.

[29] G. Makransky, T. S. Terkildsen, and R. E. Mayer, "Role of subjective and objective measures of cognitive processing during learning in explaining the spatial contiguity effect," *Learning and Instruction*, vol. 61, pp. 23–34, 2019.

[30] Z. Liu, Y.-F. Li, L.-P. He, Y.-J. Yang, and H.-Z. Huang, "A new fault tree analysis approach based on imprecise reliability model," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 228, no. 4, pp. 371–381, 2014.

[31] E. Hollnagel, *Cognitive Reliability and Error Analysis Method*, Elsevier, Amsterdam, Netherlands, 1998.

[32] K. Tyagi, A. Sharma, A. Sharma et al., "Ranking of components for reliability estimation of CBSS using fuzzy TOPSIS," *International Journal of System Assurance Engineering and Management*, vol. 7, no. S1, pp. 41–49, 2016.

[33] J. Singh, S. K. Sharma, and R. Srivastava, "AHP-Entropy based priority assessment of factors to reduce aviation fuel consumption," *International Journal of System Assurance Engineering and Management*, vol. 10, no. 2, pp. 212–227, 2019.

[34] G. D'Amico, G. D. Biase, and R. Manca, "Decomposition of the population dynamic Theil's entropy and its application to four European countries," *Hitotsubashi Journal of Economics*, vol. 55, no. 2, pp. 229–239, 2014.

[35] Z. Liu, X. Liu, H. Huang et al., "A new inherent reliability modeling and analysis method based on imprecise Dirichlet model for machine tool spindle," *Annals of Operations Research*, 2019.

[36] A. Gearhart, D. T. Booth, K. Sedivec, and C. Schauer, "Use of Kendall's coefficient of concordance to assess agreement among observers of very high resolution imagery," *Geocarto International*, vol. 28, no. 6, pp. 517–526, 2013.

[37] P. Legendre, "Species associations: the Kendall coefficient of concordance revisited," *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 10, no. 2, pp. 226–245, 2005.

[38] T. L. Saaty, "Decision-making with the AHP: why is the principal eigenvector necessary," *European Journal of Operational Research*, vol. 145, no. 1, pp. 85–91, 2003.

[39] A. R. M. Pinheiro, E. A. P. Do Prado, M. R. Martins, J. S. Gomes, and E. L. Droguett, "Neuropsychological aspects observed in a nuclear plant simulator and its relation to human reliability analysis," *Advances in Intelligent Systems and Computing*, Springer International Publishing, Berlin, Germany, pp. 145–156, 2016.