



Research article

Statistical correlation coefficients for single-valued neutrosophic sets and their applications in medical diagnosis

Xiaoyan Zhou^{1,*}, Mingwei Lin², and Weiwei Wang³

¹ Public Education Department, Fuzhou Polytechnic, Fuzhou, Fujian 350108, China

² College of Computer and Cyber Security, Fujian Normal University, Fuzhou, Fujian 350117, China

³ School of Mathematics and Statistics, Fujian Normal University, Fuzhou, Fujian 350117, China

* **Correspondence:** Email: zxyfzy2003@126.com.

Abstract: The concept of single-valued neutrosophic sets (SVNSs) is considered as an attractive tool for dealing with highly ambiguous and uncertain information. The correlation coefficient of SVNSs acts as an important measure in the single-valued neutrosophic set theory and it has been applied in various fields, such as the pattern recognition, medical diagnosis, decision-making and also clustering analysis. To alleviate the weakness of the existing correlation coefficients, a novel statistical correlation coefficient is put forward to measure the degree of correlation between two SVNSs. This statistical correlation coefficient is developed based on the variance and covariance of SVNSs and its value is between -1 and 1 . When solving the multicriteria decision making problems, the criteria show different weight values. To consider the weight information of multiple criteria, the weighted statistical correlation coefficient is developed for SVNSs. Afterwards, two numerical examples are given to show the effectiveness of the proposed statistical correlation coefficient in the pattern recognition, which can accurately classify unknown patterns into known patterns. Finally, the feasibility and practicability of the proposed correlation coefficient formula are illustrated by a practical multiple attribute decision making problem of traditional Chinese medicine diagnosis. The comparative results show that the proposed correlation coefficient formula is rational and effective.

Keywords: single-valued neutrosophic set; correlation coefficient; pattern recognition; traditional Chinese medicine diagnosis; COVID-19

Mathematics Subject Classification: 03E72, 62A86, 68T35, 90B50

1. Introduction

Medical diagnosis is of great significance to doctors to judge the situations of patients correctly. Due to the complexity of the illness, doctors are often not able to determine the specific illness of a patient based on the symptoms. There exists uncertainty during the judgment of doctors. Fuzzy sets have a great advantage in representing the uncertainty of information. In 1965, Zadeh [1] proposed a method of describing the fuzzy phenomena in mathematics, which is called the fuzzy set theory. The fuzzy set is a vital definition, which can model the uncertain, imprecise and vague information in a quantitative form. In the fuzzy set, each element is represented as a membership value that belongs to the interval $[0,1]$ [2]. However, the non-membership value (NMV) is not directly represented in the fuzzy sets [3–5]. The NMV of each fuzzy number within the fuzzy sets is considered to be the result of $(1-\text{membership value})$. In this case, the hesitant degree is ignored. To accurately describe the positive degree, hesitant degree and negative degree of fuzzy information, the concept of intuitionistic fuzzy sets (IFSs) was designed by Atanassov [6]. In an IFS, the membership value (MV) μ_p and NMV \mathcal{G}_p satisfy the following conditions: $\mu_p \in [0,1]$, $\mathcal{G}_p \in [0,1]$ and $\mu_p + \mathcal{G}_p \in [0,1]$ [7,8]. Wang et al. [9] pointed out that IFSs cannot express the indeterminate information and inconsistent information in an explicit way. The neutrosophic sets (NSs) [10,11] are a general fuzzy information modeling framework for generalizing fuzzy sets [11], IFSs, picture fuzzy sets [12–14], etc. In each NS, the truth-membership value (TV), indeterminacy-membership value (IV) and falsity-membership value (FV) are non-standard subsets of $]0-,1+[$. However, NSs should be specified. Otherwise, they cannot be directly applied in real applications. In this case, a special instantiation of neutrosophic sets, called single valued NS (SVNS), was proposed by Wang et al. [9]. At present, SVNS are widely used in the medical diagnostics. Luo et al. [15] proposed a novel distance between single-valued neutrosophic sets based on the matrix norm, and applied the proposed method to pattern recognition and medical diagnoses. Hanna et al. [16] proposed a risk classification model for cardiac patients based on the theory of neutrosophic sets and compared the proposed method with other commonly used models. Hassan et al. [17] used the neutrosophic sets to deal with the system deadlock problem in the use of electronic medical records in hospitals and achieved good results.

Similar to NSs, each single valued neutrosophic number (SVNN) consists of three components, TV, IV and FV and these three components are independent of each other [18]. Different from NSs, TV, IV and FV in SVNSs belongs to the unit interval $[0,1]$ [19–21]. Due to the complex information representation of SVNSs, various measures and decision-making methods have been proposed for handling the SVNSs [22,23]. The correlation coefficient is an important measurement metric that is used to quantify the correlation between two objects [24]. This measurement metric has been widely applied into the fields of pattern recognition [25,26], decision-making methods [27–29] and medical diagnosis [30,31]. Because of its valuable use, it has been used to measure the correlation between fuzzy variables [32]. It has also been extended to measure the correlation between IFSs [33]. However, the complex information representation of SVNSs is different from fuzzy sets and IFSs. Ye [34] first developed the correlation coefficient calculation formulas for SVNSs, and applied them for handling decision problems. Ye [35] further improved the correlation coefficient calculation formulas and also used them to improve the decision-making methods. Meng et al. [36] introduced SVNSs to represent the fuzzy information, calculated the attribute weights by the Demental-Analytical Network Process (DANP) method, and then used the evaluation based on distance from average solution (EDAS) method to rank the alternatives. Song et al. [37] used the feature of neutrosophic sets can describe the uncertainty of the problem well and used it in image segmentation algorithm. Gou et al. [38] used the

single-valued neutrosophic sets to represent trapezoidal fuzzy numbers and combined with TOPSIS method for fuzzy risk analysis.

However, through the numerical case analysis, it is known that these correlation coefficient formulas of SVNNSs [34,35] have a value range of $[0, 1]$. Meanwhile, by looking up the literatures, we also found that most of the existing correlation coefficients of SVNNSs ignore the negative correlation of two SVNNSs.

To overcome the drawbacks of these existing studies, a novel SVNNS correlation coefficient is proposed in this paper. The contributions of this paper are summarized as follows:

(1) Several counter-examples are used to analyze the drawbacks of the existing correlation coefficients of SVNNSs. To overcome these drawbacks, a novel SVNNSs correlation coefficient is proposed.

(2) Two application cases related to pattern recognition are used to show that the unknown pattern can be properly classified into the known pattern when the proposed SVNNSs correlation coefficient is used.

(3) A TCM medical diagnosis case is introduced to compare our proposed correlation coefficient with the distance and similarity of SVNNSs, the results show that the proposed correlation coefficient formula can accurately judge the patient's condition.

The rest of this paper is organized as follows. In Section 2, we present the basic knowledge about IFSs, SVNNSs and analyze the drawbacks of the existing correlation coefficients in the SVNNS environment. Section 3, a novel correlation coefficient formula of SVNNSs is proposed and its characteristics are discussed as well. Section 4 shows the application of our proposed correlation coefficient in pattern recognition. In Section 5, a TCM medical diagnosis case is provided to show the effectiveness of the proposed correlation coefficient. Finally, some valuable conclusions are summarized in Section 6.

2. Preliminaries

In this section, we will briefly describe the basic concepts and mathematical forms of IFSs and SVNNSs. Then we will analyze two existing correlation coefficient formulas about SVNNSs. Finally, two examples will be given to illustrate their limitation.

2.1. Intuitionistic fuzzy sets

IFSs [6] use pairs of MV and NMV to describe the uncertain information. The mathematical form of an IFS is described as follows.

Definition 2.1 [6]. Let $U = \{u_1, u_2, \dots, u_n\}$ be a finite universe of discourse, then an IFS P on U is defined as

$$P = \{(\mu_p(u), \mathcal{G}_p(u)) | u \in U\} \quad (1)$$

where $\mu_p(u): U \rightarrow [0,1]$ and $\mathcal{G}_p(u): U \rightarrow [0,1]$ are the MV and NMV of each element $u \in U$ belonging to the set P , respectively. They satisfy that $0 \leq \mu_p \leq 1$, $0 \leq \mathcal{G}_p \leq 1$ and $0 \leq \mu_p + \mathcal{G}_p \leq 1$. The hesitant degree or indeterminacy degree can be calculated as $\pi_p = 1 - \mu_p - \mathcal{G}_p$.

2.2. Single-valued neutrosophic sets

In SVNNSs, three important parameters are used together to express the uncertain information, which are truth-membership value (TV), indeterminacy-membership value (IV) and falsity-membership value (FV).

Definition 2.2 [34]. A single-valued neutrosophic set P on a finite universe of discourse $U = \{u_1, u_2, \dots, u_n\}$ is defined as

$$P = \{(\mu_P(u), \eta_P(u), \mathcal{G}_P(u)) | u \in U\} \quad (2)$$

where the mathematical symbols $\mu_P(u)$, $\mathcal{G}_P(u)$ and $\eta_P(u)$ denote the TV, FV and IV of each element $u \in U$ belonging to the set P , respectively. They satisfy the conditions: $0 \leq \mu_P(u), \eta_P(u), \mathcal{G}_P(u) \leq 1$ and $0 \leq \mu_P(u) + \eta_P(u) + \mathcal{G}_P(u) \leq 3$.

A triplet of TV, FV and IV of each element, denoted as $(\mu_P(u), \eta_P(u), \mathcal{G}_P(u))$, is usually called as a single-valued neutrosophic number (SVNN) [34,35]. For simplification of operations, it is usually denoted as $(\mu_P, \eta_P, \mathcal{G}_P)$.

2.3. The existing correlation coefficients for SVNNS

Definition 2.3 [34]. Let $P = \{(\mu_{P_i}, \eta_{P_i}, \mathcal{G}_{P_i}) | u_i \in U\}$ and $S = \{(\mu_{S_i}, \eta_{S_i}, \mathcal{G}_{S_i}) | u_i \in U\}$ be any two SVNNSs on the universe of discourse $U = \{u_1, u_2, \dots, u_n\}$, then the correlation coefficient between P and S was defined by Ye [34] as the following form:

$$K(P, S) = \frac{\sum_{i=1}^n (\mu_{P_i} \mu_{S_i} + \eta_{P_i} \eta_{S_i} + \mathcal{G}_{P_i} \mathcal{G}_{S_i})}{\sqrt{\sum_{i=1}^n [(\mu_{P_i})^2 + (\eta_{P_i})^2 + (\mathcal{G}_{P_i})^2]} \sqrt{\sum_{i=1}^n [(\mu_{S_i})^2 + (\eta_{S_i})^2 + (\mathcal{G}_{S_i})^2]}} \quad (3)$$

Ye [34] proved that the formula (3) satisfies the following properties:

- (1) $K(P, S) = K(S, P)$;
- (2) $K(P, S) = 1$ if $P = S$;
- (3) $0 \leq K(P, S) \leq 1$.

By observing the formula (3) and its properties, it can be easily found that the formula (3) proposed by Ye [34] cannot measure the negative correlation between two SVNNSs. To show the implementation process of the formula, an example about the calculation of the correlation coefficient between two SVNNSs is given below:

Example 2.1. Let $P = \{(0.4, 0.3, 0.1), (0.5, 0.3, 0.2), (0.4, 0.3, 0.0)\}$ and $S = \{(0.1, 0.3, 0.4), (0.2, 0.3, 0.5), (0.0, 0.3, 0.4)\}$ be two SVNNSs defined in the set $U = \{u_1, u_2, u_3\}$. If the formula (3) proposed by Ye [34] is used to measure the correlation coefficient between P and S , then we can obtain the result $K(P, S) = 0.6180$.

Definition 2.4 [35]. Let $P = \{(\mu_{P_i}, \eta_{P_i}, \mathcal{G}_{P_i}) | u_i \in U\}$ and $S = \{(\mu_{S_i}, \eta_{S_i}, \mathcal{G}_{S_i}) | u_i \in U\}$ be any two

SVNSs on the universe of discourse $U = \{u_1, u_2, \dots, u_n\}$, then the correlation coefficient between the SVNSs P and S proposed by Ye [35] was defined as the following form:

$$M(P, S) = \frac{1}{3n} \sum_{i=1}^n [\alpha_i (1 - \Delta\mu_i) + \beta_i (1 - \Delta\eta_i) + \chi_i (1 - \Delta\varrho_i)] \quad (4)$$

where

$$\alpha_i = \frac{3 - \Delta\mu_i - \Delta\mu_{\max}}{3 - \Delta\mu_{\min} - \Delta\mu_{\max}}, \quad \beta_i = \frac{3 - \Delta\eta_i - \Delta\eta_{\max}}{3 - \Delta\eta_{\min} - \Delta\eta_{\max}}, \quad \chi_i = \frac{3 - \Delta\varrho_i - \Delta\varrho_{\max}}{3 - \Delta\varrho_{\min} - \Delta\varrho_{\max}},$$

$$\Delta\mu_i = |\mu_{P_i} - \mu_{S_i}|, \quad \Delta\mu_{\min} = \min_i |\mu_{P_i} - \mu_{S_i}|, \quad \Delta\mu_{\max} = \max_i |\mu_{P_i} - \mu_{S_i}|,$$

$$\Delta\eta_i = |\eta_{P_i} - \eta_{S_i}|, \quad \Delta\eta_{\min} = \min_i |\eta_{P_i} - \eta_{S_i}|, \quad \Delta\eta_{\max} = \max_i |\eta_{P_i} - \eta_{S_i}|,$$

$$\Delta\varrho_i = |\varrho_{P_i} - \varrho_{S_i}|, \quad \Delta\varrho_{\min} = \min_i |\varrho_{P_i} - \varrho_{S_i}|, \quad \Delta\varrho_{\max} = \max_i |\varrho_{P_i} - \varrho_{S_i}|.$$

Ye [35] proved that the formula (4) also meets the following properties:

- (1) $M(P, S) = M(S, P)$;
- (2) $M(P, S) = 1$ if $P = S$;
- (3) $0 \leq M(P, S) \leq 1$.

By observing the formula (4), it can be easily seen that the formula (4) proposed by Ye [35] also cannot measure the negative correlation between two SVNSs.

Example 2.2. Let $P = \{(0.4, 0.3, 0.1), (0.5, 0.3, 0.2), (0.4, 0.3, 0.0)\}$ and $S = \{(0.1, 0.3, 0.4), (0.2, 0.3, 0.5), (0.0, 0.3, 0.4)\}$ be two SVNSs defined in the set $U = \{u_1, u_2, u_3\}$. If the formula (4) proposed by Ye [35] is used to measure the correlation coefficient between P and S, then we can obtain the result $M(P, S) = 0.7504$.

In summary, these two correlation coefficient formulas between SVNSs proposed by Ye [34] and Ye [35] cannot measure the negative correlation between two SVNSs. To solve this problem, we will propose a novel formula of correlation coefficient between two SVNSs, the value of which falls into the range of -1 to 1 .

3. Statistical correlation coefficients between two SVNSs

In this section, the statistical theory is used to develop some novel correlation coefficients for SVNSs, and the proposed statistical correlation coefficients are compared with the existing correlation coefficients to show the usefulness of the proposed statistical correlation coefficients. Before proposing the novel formulas of correlation coefficient between two SVNSs, it is necessary to develop some basic concepts based on the statistical knowledge for SVNSs.

First, the variance of an SVNS P is developed by introducing the variance concept from statistics.

Definition 3.1. Let $P = \{(\mu_{P_i}, \eta_{P_i}, \varrho_{P_i}) | u_i \in U\}$ be an SVNS on the universe of discourse $U = \{u_1, u_2, \dots, u_n\}$, then the variance of P is statistically defined as:

$$D(P) = \frac{1}{n} \sum_{i=1}^n d_i^2(P) \quad (5)$$

where the term $d_i(P)$ is calculated as:

$$d_i(P) = \left(\mu_{p_i}^2 - \overline{\mu_p}^2 \right) - \left(\eta_{p_i}^2 - \overline{\eta_p}^2 \right) - \left(\mathcal{G}_{p_i}^2 - \overline{\mathcal{G}_p}^2 \right)$$

with

$$\overline{\mu_p} = \frac{1}{n} \sum_{i=1}^n \mu_{p_i}, \quad \overline{\eta_p} = \frac{1}{n} \sum_{i=1}^n \eta_{p_i}, \quad \overline{\mathcal{G}_p} = \frac{1}{n} \sum_{i=1}^n \mathcal{G}_{p_i}, \quad \forall_i = 1, 2, \dots, n.$$

Next, the covariance formula between two SVNNS is developed based on the conventional statistical viewpoint of covariance of two sets.

Definition 3.2. Let $P = \{(\mu_{p_i}, \eta_{p_i}, \mathcal{G}_{p_i}) | u_i \in U\}$ and $S = \{(\mu_{s_i}, \eta_{s_i}, \mathcal{G}_{s_i}) | u_i \in U\}$ be any two SVNNS on the universe of discourse $U = \{u_1, u_2, \dots, u_n\}$, then the covariance of SVNNS P and S is statistically defined as:

$$COV(P, S) = \frac{1}{n} \sum_{i=1}^n d_i(P) d_i(S) \quad (6)$$

where the terms $d_i(P)$ and $d_i(S)$ are calculated as:

$$d_i(P) = \left(\mu_{p_i}^2 - \overline{\mu_p}^2 \right) - \left(\eta_{p_i}^2 - \overline{\eta_p}^2 \right) - \left(\mathcal{G}_{p_i}^2 - \overline{\mathcal{G}_p}^2 \right), \quad d_i(S) = \left(\mu_{s_i}^2 - \overline{\mu_s}^2 \right) - \left(\eta_{s_i}^2 - \overline{\eta_s}^2 \right) - \left(\mathcal{G}_{s_i}^2 - \overline{\mathcal{G}_s}^2 \right).$$

Now, the novel formulas of correlation coefficient between SVNNS are developed based on the Pearson correlation coefficient in statistics.

Definition 3.3. Let $P = \{(\mu_{p_i}, \eta_{p_i}, \mathcal{G}_{p_i}) | u_i \in U\}$ and $S = \{(\mu_{s_i}, \eta_{s_i}, \mathcal{G}_{s_i}) | u_i \in U\}$ be any two SVNNS on the universe of discourse $U = \{u_1, u_2, \dots, u_n\}$, then the statistical correlation coefficient between SVNNS P and S is defined as:

$$\rho(P, S) = \frac{COV(P, S)}{[D(P)D(S)]^{\frac{1}{2}}} = \frac{\sum_{i=1}^n d_i(P) d_i(S)}{\sqrt{\sum_{i=1}^n d_i^2(P) \sum_{i=1}^n d_i^2(S)}} \quad (7)$$

Remark 3.1. The value of the proposed statistical correlation coefficient formula falls into the range $[-1, 1]$, while the value range of the existing correlation coefficient formulas is $[0, 1]$. As we know, the correlation coefficient is used to measure the linear relationship between any two variables or subsets. Both of positive correlation and negative correlation have practical significance, and it is unreasonable to give up any part of them. Thus, the proposed statistical correlation coefficient formula obtains more reasonable values.

Properties 3.1. Given SVNNS $P = \{(\mu_{p_i}, \eta_{p_i}, \mathcal{G}_{p_i}) | u_i \in U\}$ and $S = \{(\mu_{s_i}, \eta_{s_i}, \mathcal{G}_{s_i}) | u_i \in U\}$, then we

have

- (1) $\rho(P, S) = \rho(S, P)$;
- (2) $-1 \leq \rho(P, S) \leq 1$.

Proof. (1) Straightforward.

(2) According to the Cauchy-Schwarz inequality $\left(\sum_{k=1}^m u_k v_k\right)^2 \leq \left(\sum_{k=1}^m u_k^2\right)\left(\sum_{k=1}^m v_k^2\right)$ with $u = (u_1, u_2, \dots, u_m)$ and $v = (v_1, v_2, \dots, v_m)$, then $(COV(P, S))^2 = \frac{1}{n^2} \left\{ \sum_{i=1}^n d_i(P) d_i(S) \right\}^2 \leq \frac{1}{n} \sum_{i=1}^n d_i^2(P) \times \frac{1}{n} \sum_{i=1}^n d_i^2(S) = D(P)D(S)$.

Then, we can obtain that:

$$|COV(P, S)| \leq (D(P))^{\frac{1}{2}} (D(S))^{\frac{1}{2}} \quad \text{and} \quad -(D(P))^{\frac{1}{2}} (D(S))^{\frac{1}{2}} \leq COV(P, S) \leq (D(P))^{\frac{1}{2}} (D(S))^{\frac{1}{2}}.$$

Finally, we have
$$-1 \leq \rho(P, S) = \frac{COV(P, S)}{[D(P)D(S)]^{\frac{1}{2}}} \leq 1.$$

Example 3.1. Let $P = \{(0.4, 0.3, 0.1), (0.5, 0.3, 0.2), (0.4, 0.3, 0.0)\}$ and $S = \{(0.1, 0.3, 0.4), (0.2, 0.3, 0.5), (0.0, 0.3, 0.4)\}$ be two SVNSSs, then using the formula (7) to calculate the correlation coefficient value between P and S can obtain $\rho(P, S) = -1$.

By comparing the results of Example 2.1, Example 2.2 and Example 3.1, it can be found that the result of the proposed statistical correlation coefficient is more reasonable, and the proposed statistical correlation coefficient formula can obtain negative correlation between two SVNSSs.

In some practical problems, the elements in the set $U = \{u_1, u_2, \dots, u_n\}$ shows different importance. For example, the criteria in the multicriteria decision-making problems usually own different importance. To let the proposed formula (7) of the statistical correlation coefficient be applied to solve these practical problems, the proposed formula (7) is extended by considering the importance of elements, and its definition is given below.

Definition 3.4. Let $P = \{(\mu_{P_i}, \eta_{P_i}, \mathcal{G}_{P_i}) | u_i \in U\}$ and $S = \{(\mu_{S_i}, \eta_{S_i}, \mathcal{G}_{S_i}) | u_i \in U\}$ be any two SVNSSs on the universe of discourse $U = \{u_1, u_2, \dots, u_n\}$, then the weighted statistical correlation coefficient value between SVNSSs P and S is calculated as:

$$\rho_{\omega}(P, S) = \frac{COV_{\omega}(P, S)}{[D_{\omega}(P)D_{\omega}(S)]^{\frac{1}{2}}} = \frac{\sum_{i=1}^n \omega_i d_i(P) d_i(S)}{\sqrt{\sum_{i=1}^n \omega_i d_i^2(P) \sum_{i=1}^n \omega_i d_i^2(S)}} \quad (8)$$

where the term ω_i denotes the importance of the element u_i in U , and $COV_{\omega}(P, S) = \frac{1}{n} \sum_{i=1}^n \omega_i d_i(P) d_i(S)$, $D_{\omega}(P) = \frac{1}{n} \sum_{i=1}^n \omega_i d_i^2(P)$, $D_{\omega}(S) = \frac{1}{n} \sum_{i=1}^n \omega_i d_i^2(S)$.

Properties 3.2. Given SVNSs $P = \{(\mu_{P_i}, \eta_{P_i}, \mathcal{G}_{P_i}) | u_i \in U\}$ and $S = \{(\mu_{S_i}, \eta_{S_i}, \mathcal{G}_{S_i}) | u_i \in U\}$, then we have

- (1) $\rho_\omega(P, S) = \rho_\omega(S, P)$;
- (2) $-1 \leq \rho_\omega(P, S) \leq 1$.

Proof. (1) Straightforward.

(2) According to the Cauchy-Schwarz inequality and the proposed formula (8), the following inequality can be obtained as

$$COV_\omega(P, S)^2 = \frac{1}{n^2} \left\{ \sum_{i=1}^n \omega_i d_i(P) d_i(S) \right\}^2 \leq \frac{1}{n} \sum_{i=1}^n \omega_i d_i^2(P) \times \frac{1}{n} \sum_{i=1}^n \omega_i d_i^2(S) = D_\omega(P) D_\omega(S).$$

Then, we can obtain that:

$$|COV_\omega(P, S)| \leq (D_\omega(P))^{1/2} (D_\omega(S))^{1/2} \quad \text{and} \quad -(D_\omega(P))^{1/2} (D_\omega(S))^{1/2} \leq COV_\omega(P, S) \leq (D_\omega(P))^{1/2} (D_\omega(S))^{1/2}.$$

Hence, we have
$$-1 \leq \rho_\omega(P, S) = \frac{COV_\omega(P, S)}{[D_\omega(P) D_\omega(S)]^{1/2}} \leq 1.$$

In the next section, we will apply the proposed statistical correlation coefficients in the form of formulas (7) and (8) to solve the problems of pattern recognition and medical diagnosis.

4. Applications of the proposed statistical correlation coefficients

To show the usefulness and superiority of the proposed statistical correlation coefficients in terms of the formulas (7) and (8), the proposed statistical correlation coefficient formulas are applied to solve two practical problems, which are pattern recognition and medical diagnosis, respectively.

4.1. Application of the statistical correlation coefficient in pattern recognition

In the practical pattern recognition problem, an unknown pattern is classified into some known patterns with the help of various information measures such as the distance measure, divergence measure, similarity measure, correlation coefficient, etc. Here, the proposed statistical correlation coefficient formula is used for solving the pattern recognition problem. To verify the usefulness and superiority of the proposed statistical correlation coefficient formula, the proposed statistical correlation coefficient is compared with the existing information measures.

First, a classical pattern recognition problem under the SVNS environment is formulated as follows:

Problem formulation: Suppose that $\{P_1, P_2, \dots, P_n\}$ are a series of known patterns that are characterized by the following SVNS $P_j = \{(\mu_{P_j}, \eta_{P_j}, \mathcal{G}_{P_j}) | u_i \in U\}$. Let $R = \{(\mu_R, \eta_R, \mathcal{G}_R) | u_i \in U\}$ be an unknown pattern, then this pattern recognition problem is how to classify the unknown pattern R to one of the known patterns P_j ($j = 1, 2, \dots, n$).

There exist various information measures that can be used to solve this pattern recognition problem. In this section, we use three kinds of information measures to solve the pattern recognition

problem, which are correlation coefficient, distance measure and similarity measure, respectively. The methods based on these information measures are given as follows:

(1) Correlation coefficient method

Let $C(P_j, R)$ be the correlation coefficient value between SVNS R and $P_j (j=1, 2, \dots, n)$, then R is considered to be in the same pattern of P_{j^*} , where $j^* = \operatorname{argmax} \{C(P_j, R)\}$, $j=1, 2, \dots, n$.

(2) Distance measure method

Let $d(P_j, R)$ be the distance value between SVNS R and $P_j (j=1, 2, \dots, n)$, then R is considered to be in the same pattern of P_{j^*} , where $j^* = \operatorname{argmin} \{d(P_j, R)\}$, $j=1, 2, \dots, n$.

(3) Similarity measure method

Let $S(P_j, R)$ be the similarity value between SVNS R and $P_j (j=1, 2, \dots, n)$, then R is considered to be in the same pattern of P_{j^*} , where $j^* = \operatorname{argmax} \{S(P_j, R)\}$, $j=1, 2, \dots, n$.

To compare the proposed statistical correlation coefficient with the existing information measures in the pattern recognition problem, we introduce some information measures for SVNSs as follows:

Definition 4.1 [39]. Let $P = \{(\mu_{P_i}, \eta_{P_i}, \mathcal{G}_{P_i}) | u_i \in U\}$ and $S = \{(\mu_{S_i}, \eta_{S_i}, \mathcal{G}_{S_i}) | u_i \in U\}$ be two SVNSs on the universe of discourse $U = \{u_1, u_2, \dots, u_n\}$, then the Hamming distance between SVNSs P and S are defined as:

$$d_1(P, S) = \frac{1}{2n} \sum_{i=1}^n \{|\mu_{P_i} - \mu_{S_i}| + |\eta_{P_i} - \eta_{S_i}| + |\mathcal{G}_{P_i} - \mathcal{G}_{S_i}|\} \quad (9)$$

Definition 4.2 [40]. Let $P = \{(\mu_{P_i}, \eta_{P_i}, \mathcal{G}_{P_i}) | u_i \in U\}$ and $S = \{(\mu_{S_i}, \eta_{S_i}, \mathcal{G}_{S_i}) | u_i \in U\}$ be two SVNSs on the universe of discourse $U = \{u_1, u_2, \dots, u_n\}$, then the Euclidean distance between SVNSs P and S are defined as:

$$d_2(P, S) = \sqrt{\frac{1}{2n} \sum_{i=1}^n \{(\mu_{P_i} - \mu_{S_i})^2 + (\eta_{P_i} - \eta_{S_i})^2 + (\mathcal{G}_{P_i} - \mathcal{G}_{S_i})^2\}} \quad (10)$$

Definition 4.3 [41]. Let $P = \{(\mu_{P_i}, \eta_{P_i}, \mathcal{G}_{P_i}) | u_i \in U\}$ and $S = \{(\mu_{S_i}, \eta_{S_i}, \mathcal{G}_{S_i}) | u_i \in U\}$ be two SVNSs on the universe of discourse $U = \{u_1, u_2, \dots, u_n\}$, then the Dice similarity between SVNSs P and S are defined as:

$$S_1(P, S) = \frac{1}{n} \sum_{i=1}^n \left[\frac{2(\mu_{P_i} \mu_{S_i} + \eta_{P_i} \eta_{S_i} + \mathcal{G}_{P_i} \mathcal{G}_{S_i})}{((\mu_{P_i}^2 + \eta_{P_i}^2 + \mathcal{G}_{P_i}^2) + (\mu_{S_i}^2 + \eta_{S_i}^2 + \mathcal{G}_{S_i}^2))} \right] \quad (11)$$

Definition 4.4 [41]. Let $P = \{(\mu_{P_i}, \eta_{P_i}, \mathcal{G}_{P_i}) | u_i \in U\}$ and $S = \{(\mu_{S_i}, \eta_{S_i}, \mathcal{G}_{S_i}) | u_i \in U\}$ be two SVNSs on the universe of discourse $U = \{u_1, u_2, \dots, u_n\}$, then the cosine similarity between SVNSs P and S are defined as:

$$S_2(P, S) = \frac{1}{n} \sum_{i=1}^n \frac{\mu_{P_i} \mu_{S_i} + \eta_{P_i} \eta_{S_i} + \mathcal{G}_{P_i} \mathcal{G}_{S_i}}{\left(\sqrt{\mu_{P_i}^2 + \eta_{P_i}^2 + \mathcal{G}_{P_i}^2} \times \sqrt{\mu_{S_i}^2 + \eta_{S_i}^2 + \mathcal{G}_{S_i}^2} \right)} \quad (12)$$

Here, we give a specific example about solving the pattern recognition problem as follows:

Example 4.1. There exist three known patterns P_1 , P_2 and P_3 that are modelled as the following SVNNSs:

$$\begin{aligned} P_1 &= \{(0.4, 0.3, 0.0), (0.5, 0.1, 0.5), (0.3, 0.0, 0.5), (0.0, 0.6, 0.4), (0.5, 0.1, 0.3)\}, \\ P_2 &= \{(0.7, 0.2, 0.2), (0.3, 0.7, 0.1), (0.5, 0.2, 0.5), (0.2, 0.3, 0.7), (0.0, 0.1, 0.3)\}, \\ P_3 &= \{(0.1, 0.1, 0.2), (0.33, 0.42, 0.55), (0.6, 0.5, 0.6), (0.8, 0.2, 0.2), (0.9, 0.1, 0.0)\}. \end{aligned}$$

Let R be an unknown pattern that is modelled as the following SVNNS:

$$R = \{(0.5, 0.5, 0.5), (0.4, 0.1, 0.4), (0.7, 0.0, 0.7), (0.1, 0.9, 0.1), (0.3, 0.8, 0.3)\}.$$

We use the formulas (3), (4), (7), (9), (10), (11) and (12) to calculate the relationship value between the unknown pattern and each known pattern, and then we can classify the unknown pattern R into one of the known patterns P_j ($j=1, 2, 3$). The results of these information measures are given in Table 1.

Table 1. Results of different information measures regarding Example 4.1.

	(P_1, R)	(P_2, R)	(P_3, R)	Results
$K(P_i, R)$	0.8212	0.6708	0.6030	P_1
$M(P_i, R)$	0.7440	0.6398	0.5980	P_1
$\rho(P_i, R)$	0.4143	0.3248	-0.8804	P_1
$d_1(P_i, R)$	0.2133	0.3133	0.3627	P_1
$d_2(P_i, R)$	1.2400	2.1100	2.7398	P_1
$S_1(P_i, R)$	0.7862	0.6241	0.6004	P_1
$S_2(P_i, R)$	0.1642	0.1341	0.1206	P_1

In Table 1, it can be easily found that all these information measures including our proposed statistical correlation coefficient achieve the same results and then classify the unknown pattern R to P_1 . It shows that the classification result that is obtained by our proposed statistical correlation coefficient is consistent with those that were obtained by the existing information measures.

Next, we consider another example about the pattern recognition problem to show the efficiency of our proposed statistical correlation coefficient in the counter-intuitive situation.

Example 4.2. There exist three known patterns P_1 , P_2 and P_3 that are modeled as the following SVNNSs:

$$P_1 = \{(0.4, 0.4, 0.4), (0.3, 0.3, 0.2), (0.3, 0.3, 0.0), (0.3, 0.0, 0.2), (0.3, 0.1, 0.1)\},$$

$$P_2 = \{(0.7, 0.1, 0.1), (0.2, 0.3, 0.4), (0.2, 0.1, 0.5), (0.1, 0.5, 0.2), (0.3, 0.3, 0.3)\},$$

$$P_3 = \{(0.2, 0.3, 0.4), (0.4, 0.3, 0.2), (0.3, 0.4, 0.2), (0.2, 0.5, 0.3), (0.5, 0.3, 0.1)\}.$$

Let R be an unknown pattern given in terms of a SVNS as

$$R = \{(0.1, 0.2, 0.1), (0.3, 0.4, 0.2), (0.4, 0.3, 0.2), (0.7, 0.1, 0.0), (0.4, 0.2, 0.2)\}.$$

We use the formulas (3), (4), (7), (9), (10), (11), (12) to calculate the relationship value between the unknown pattern and each known pattern, and then we can classify the unknown pattern R into one of the known patterns P_j ($j = 1, 2, 3$). The results of these information measures are given in Table 2.

Table 2. Results of different information measures regarding Example 4.2.

	(P_1, R)	(P_2, R)	(P_3, R)	Results
$K(P_i, R)$	0.7963	0.6126	0.7739	P_1
$M(P_i, R)$	0.8116	0.7883	0.8171	P_3
$\rho(P_i, R)$	0.4673	-0.4024	-0.5474	P_1
$d_1(P_i, R)$	0.1467	0.2200	0.1600	P_1
$d_2(P_i, R)$	0.5200	1.1900	0.6800	P_1
$S_1(P_i, R)$	0.8095	0.6401	0.7995	P_1
$S_2(P_i, R)$	0.1593	0.1225	0.1548	P_1

From Table 2, we can find that the classification result obtained by the formula (4) is different from the classification results obtained by others including the formula (7). The formulas (3), (7), (9), (10), (11) and (12) can accurately classify the unknown pattern R to P_1 . By observing the data in Example 4.2 and the classification results in Table 2, it is easy to find that the formula (4) has erroneous judgement. The formula (7) of the proposed statistical correlation coefficient not only can calculate the negative correlation value of two patterns, but also obtain more robust classification result.

4.2. Application of the statistical correlation coefficient in medical diagnosis

Medical diagnosis is a process of finding the diseased part, judging the degree of the disease and then determining the disease according to the symptoms of the patient when their body is in an abnormal state.

This section attempts to validate the rationality and the feasibility of the proposed statistical correlation coefficient during the diagnosis process of patients with COVID-19 by Chinese medicine. We have collected diagnosis and treatment data of suspected patients with COVID-19 from the relevant medical departments. In the following part, we first give the problem background. Next, we will use various information measures to diagnose the suspected cases, and also compare these information measures to show the effectiveness and practicability of the proposed statistical correlation coefficient.

4.2.1. Problem background

As we know, the root of Tradition Chinese Medicine (also referred to as TCM) dates back more than 2000 years. Tradition Chinese Medicine (TCM) is a complete medical system that has been used to diagnose, treat and prevent illnesses for more than 2000 years. Nowadays, TCM is an indispensable part of Chinese culture which has made great contributions to the prosperity of China. Both TCM and western medicine are being used to cure people all around the world. The TCM, comparing with western medicine, with its unique diagnostic methods, long history and remarkable effects, have been used to treat cancer and other serious diseases. Unlike the western medicine, the TCM has many advantages. One of the most significant is that it has fewer side effects. According to the survey data, TCM is used by 75% of the areas in China and has been very effective in the treatment of diseases such as diabetes, liver cancer, tumors and bone fracture and so on. Great successes have been made in many areas through TCM cure. As to acute abdomen, there is no need to have an operation. All you need to do is drink a cup of Chinese herbs, while the western way takes more time and money. You may even take the risk of being infected after operation.

In particular, TCM has achieved gratifying results in the diagnosis and treatment of COVID-19. This also makes more and more people pay more attention to TCM. From ancient times to the present, there have been theories of yin yang philosophy in China. TCM is based on a belief in yin and yang, which is defined as opposing energies, such as earth and heaven, winter and summer and happiness and sadness. It can just be described by truth membership and falsehood non-membership in the fuzzy set. In terms of diagnosing, TCM diagnostics is to grasp the human system as a whole emerged as the goal established, the information obtained through the four clinics of the look and smell are the characteristics of information emerged about the body as a whole. It fully shows that in the diagnosis process of TCM, uncertain and vague information will be produced. For this part of the information, we use the indeterminacy-membership to describe it. Thus, we can completely introduce the related concepts of single-value neutrosophic sets in TCM diagnosis and treatment.

As the symptoms of patients with COVID-19 are similar to those of traditional pneumonia and viral influenza, a large number of suspected cases will make the diagnosis of doctors extremely difficult. Once the doctor makes a misjudgment, it will delay the opportunity for treatment and even endanger the patients' life. The notion of similarity measure, divergence measure, correlation measure and distance measure under the single-value neutrosophic set environment has played a key role in the medical diagnosis problems. Here we use our proposed statistical correlation coefficient to deal with the medical diagnosis problem. The symptom values of the diseases are assigned by one specialist doctor or by the aggregated opinion of several specialist doctors. The symptom values to a patient are assigned by the doctor dealing with the patient based on his/her expertise. Then the correlation between the symptoms of patients and the standardized symptoms of diseases is computed. The patient will be diagnosed with the disease, which has the highest value of correlation.

Now, we use our proposed statistical correlation coefficient in the medical diagnosis problem with the help of the following example.

4.2.2. Illustrative example

This application illustrates how the proposed statistical correlation coefficient can be used to address a problem of misdiagnosis in the medical diagnosis of COVID-19. The flowchart of proposed statistical correlation coefficient is shown in Figure 1.

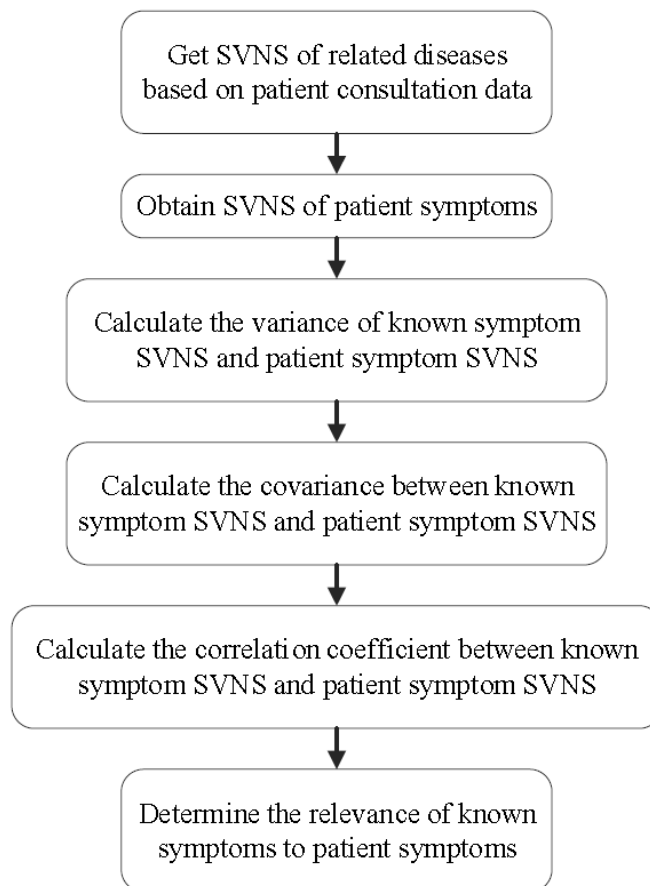


Figure 1. The flowchart of proposed statistical correlation coefficients.

Let us assume that there are four different patients $P = \{P_1, P_2, P_3, P_4\}$ and also five kinds of diseases $D = \{D_1, D_2, D_3, D_4, D_5\} = \{Malaria, SARS, Viral - influenza, COVID-19, Typhoid\}$ along with their symptoms $S = \{S_1, S_2, S_3, S_4, S_5\} = \{Fever, Pharyngalgia, Dizziness, Rhinorrhea, Dry - Cough\}$.

Example 4.3. We have collected the patients' diagnosis and treatment data from the relevant departments and organized them into the following tables. Table 3 is the dataset of symptoms of various diseases and Table 4 is the dataset of symptoms of different patients. We assume that patient P_1 is suffering from Typhoid, P_2 and P_3 are suffering from COVID-19 and P_4 is suffering from Viral Influenza. Thus, the symptoms of the patients P_j ($j = 1, 2, 3, 4$) are deduced accordingly and shown in Table 4. We use the information measures for better diagnosis.

Table 3. Symptoms of the diseases modelled as SVNSs.

	Fever	Pharyngalgia	Dizziness	Rhinorrhea	Dry Cough
Malaria	(0.8,0.1,0)	(0.2,0.4,0.35)	(0.2,0.5,0.3)	(0,0.35,0.5)	(0.1,0.5,0.3)
SARS	(0,0.1,0.8)	(0.35,0.4,0.2)	(0.3,0.5,0.2)	(0.5,0.35,0)	(0.3,0.5,0.1)
Viral influenza	(0.1,0.6,0.2)	(0.2,0.3,0.35)	(0.2,0.4,0.3)	(0.6,0.1,0.2)	(0.3,0.3,0.4)
COVID-19	(0.2,0.6,0.2)	(0.35,0.3,0.2)	(0.3,0.4,0.2)	(0.2,0.1,0.6)	(0.4,0.3,0.3)
Typhoid	(0.2,0.2,0.2)	(0.8,0.2,0.1)	(0.5,0.3,0.5)	(0.6,0.3,0.5)	(0.8,0.2,0.1)

Table 4. Symptoms of the patients modelled as SVNSSs.

	Fever	Pharyngalgia	Dizziness	Rhinorrhea	Dry Cough
P_1	(0.2,0.3,0.5)	(0.6,0.2,0.1)	(0.1,0.2,0.4)	(0.4,0.2,0.3)	(0.8,0.1,0)
P_2	(0.2,0.3,0.4)	(0.4,0.3,0.2)	(0.3,0.4,0.2)	(0.2,0.5,0.3)	(0.5,0.3,0.1)
P_3	(0.4,0.5,0.1)	(0.3,0.3,0.2)	(0.3,0.3,0)	(0.3,0,0.8)	(0.8,0.1,0.1)
P_4	(0.5,0.7,0.2)	(0.3,0.4,0.2)	(0.3,0.5,0.1)	(0.5,0.2,0.1)	(0.4,0.3,0.2)

By using the formula (3), we calculate the correlation coefficient between the patient's symptoms and disease symptoms. The result values are given in Table 5.

Table 5. Correlation coefficients between the patients' symptoms and disease symptoms using (3).

	Malaria	SARS	Viral Influenza	COVID-19	Typhoid
P_1	0.5059	0.7526	0.7291	0.7800	0.9241
P_2	0.7327	0.8589	0.7937	0.8709	0.8709
P_3	0.6491	0.5067	0.7269	0.9020	0.8094
P_4	0.7198	0.7369	0.9082	0.8708	0.7517

As shown in Table 5, it can be found that the correlation coefficient value between the symptoms of P_1 and the symptoms of Typhoid is highest. Thus, we can know that the patient P_1 is suffering from Typhoid. Similarly, it is easy to know that P_3 is suffering from COVID-19 and P_4 is suffering from Viral Influenza. However, the correlation coefficient value between the symptoms of the patient P_2 and the symptoms of COVID-19 is equal to the correlation coefficient value between the symptoms of the patient P_2 and the symptoms of Typhoid. Thus, it is difficult to diagnose the disease condition of the patient P_2 .

By using the formula (4), we calculate the correlation coefficient between the patient's symptoms and disease symptoms. The result values are given in Table 6.

Table 6. Correlation coefficients between the patients' symptoms and disease symptoms using (4).

	Malaria	SARS	Viral influenza	COVID-19	Typhoid
P_1	0.6337	0.7352	0.7446	0.7503	0.8371
P_2	0.7491	0.8225	0.8027	0.8689	0.7809
P_3	0.6851	0.6626	0.7446	0.8324	0.7473
P_4	0.7418	0.7942	0.8508	0.8520	0.7164

As shown in Table 6, it is not difficult to notice that P_1 is suffering from Typhoid, while P_2 , P_3

and P_4 are suffering from COVID-19. This is not in line with the true state of affairs, which is P_4 suffering from Viral influenza.

By using the formula (7), we calculate the correlation coefficient between the patient's symptoms and disease symptoms. The result values are given in Table 7.

Table 7. Correlation coefficients between the patients' symptoms and disease symptoms using (7).

	Malaria	SARS	Viral influenza	COVID-19	Typhoid
P_1	-0.5621	0.4184	0.1803	0.6639	0.9198
P_2	-0.3217	0.0384	-0.3823	0.8747	0.8490
P_3	-0.1009	-0.1845	-0.5223	0.7018	0.5857
P_4	-0.7114	0.8498	0.9228	0.0332	0.2011

As shown in Table 7, it is not hard to notice that P_1 is suffering from Typhoid, both P_2 and P_3 are suffering from COVID-19 and P_4 is suffering from Viral influenza. All the calculated results conform with the actual status.

By using the formula (9), we calculate the distance values between the patient's symptoms and disease symptoms. The result values are given in Table 8.

Table 8. Distance values between the patients' symptoms and disease symptoms using (9).

	Malaria	SARS	Viral influenza	COVID-19	Typhoid
P_1	0.3200	0.2333	0.2233	0.2100	0.1267
P_2	0.2200	0.1467	0.1700	0.1033	0.1800
P_3	0.2733	0.2867	0.2167	0.1300	0.2000
P_4	0.2200	0.1667	0.1233	0.1167	0.2400

As shown in Table 8, it is not difficult to notice that P_1 is suffering from Typhoid, while P_2 , P_3 and P_4 are suffering from COVID-19. This is not in line with the true state of affairs, which is P_4 suffering from Viral influenza.

Similarly, the formula (10) is used to calculate the distance values between the patient's symptoms and disease symptoms. The result values are given in Table 9.

Table 9. Distance values between the patients' symptoms and disease symptoms using (10).

	Malaria	SARS	Viral influenza	COVID-19	Typhoid
P_1	1.9850	0.9950	0.9925	0.8125	0.4300
P_2	1.0050	0.5450	0.6825	0.4325	0.7100
P_3	1.5050	2.1150	1.0925	0.4125	0.9800
P_4	1.1450	1.0750	0.3525	0.4925	1.2400

As shown in Table 9, it is easy to notice that the patient P_1 is suffering from Typhoid, both P_2 and P_3 are suffering from COVID-19 and P_4 is suffering from Viral influenza. All these results conform with the actual status.

By using the similarity formula (11), we calculate the similarity values between the patient's symptoms and disease symptoms. The result values are given in Table 10.

Table 10. Similarity values between the patients' symptoms and disease symptoms using (11).

	Malaria	SARS	Viral influenza	COVID-19	Typhoid
P_1	0.5497	0.7577	0.7497	0.7846	0.8865
P_2	0.7524	0.8652	0.8088	0.8857	0.8434
P_3	0.6811	0.5887	0.7562	0.9052	0.7742
P_4	0.7348	0.8324	0.9127	0.8710	0.7394

As shown in Table 10, it is easy to find that the patient P_1 is suffering from Typhoid, both P_2 and P_3 are suffering from COVID-19 and P_4 is suffering from Viral influenza. All these results conform with the actual status.

Similarly, the formula (12) is used to compute the similarity values between the patient's symptoms and disease symptoms. The result values are given in Table 11.

Table 11. Similarity values between the patients' symptoms and disease symptoms using (12).

	Malaria	SARS	Viral influenza	COVID-19	Typhoid
P_1	0.1012	0.1505	0.1458	0.1560	0.1848
P_2	0.1465	0.1718	0.1587	0.1742	0.1742
P_3	0.1298	0.1013	0.1454	0.1804	0.1619
P_4	0.1440	0.1474	0.1816	0.1742	0.1503

As shown in Table 11, it is not difficult to know that the patient P_1 is suffering from Typhoid. P_3 is suffering from COVID-19 and P_4 is suffering from Viral Influenza. However, it is difficult to diagnose the disease condition of the patient P_2 since the similarity value between the symptoms of the patient P_2 and the symptoms of COVID-19 is equal to the similarity value between the symptoms of the patient P_2 and the symptoms of Typhoid.

In order to compare and analyze the diagnosis results of different methods more clearly, all considered calculation results are summarized into Table 12.

Table 12. All considered results.

	P_1	P_2	P_3	P_4
Using the formula (3)	Typhoid	Unable to classify	COVID-19	Viral Influenza
Using the formula (4)	Typhoid	COVID-19	COVID-19	COVID-19
Using the formula (7) (Our proposed)	Typhoid	COVID-19	COVID-19	Viral Influenza
Using the formula (9)	Typhoid	COVID-19	COVID-19	COVID-19
Using the formula (10)	Typhoid	COVID-19	COVID-19	Viral Influenza
Using the formula (11)	Typhoid	COVID-19	COVID-19	Viral Influenza
Using the formula (12)	Typhoid	Unable to classify	COVID-19	Viral Influenza

Compared the diagnosis results of various methods with the actual situation, it is not difficult to find that the results of the formulas (4) and (9) produce misdiagnosis. Meanwhile, based on the results of formulas (3) and (12), it is unable to determine the disease condition of the patient P_2 . Using our proposed formula (7), the distance formula (10) and similarity formula (11), we can correctly diagnose the patient's disease condition.

5. Conclusions

In this paper, we propose a novel statistical correlation coefficient formula to measure the correlation values between SVNSs and also develop a weighted form of this statistical correlation coefficient formula. The value of the statistical correlation coefficients falls into the interval $[-1,1]$ which is in accordance with the range of statistical correlation coefficient between two variables or sets. Conversely, most of the existing correlation coefficient formulas of SVNSs only obtained the values in interval $[0,1]$. Some examples are given to demonstrate that the statistical correlation coefficient formula is more effective than the existing methods. In the practical application problems of pattern recognition and TCM medical diagnosis, we clearly found the superiority of the statistical correlation coefficient formula we proposed. In addition, compared the proposed statistical correlation coefficient formula with the existing correlation coefficients, distance measures and similarity measures, the statistical correlation coefficient formula we proposed can properly determine the patient's disease and reduce misjudgments. The advantages of our studies are summarized as:

(1) The value of our statistical correlation coefficients falls into the interval $[-1,1]$. The correlation coefficient formula can be used to calculate both the positive and negative correlation between single-value neutrosophic sets.

(2) In the application example of pattern recognition, our formula for calculating the correlation coefficient has been found to effectively classify the unknown pattern as the known pattern.

(3) In the application example of medical diagnosis, comparing with the distance measure methods and similarity measure methods, we can find that the correlation coefficient formula we proposed can accurately judge the patient's disease condition.

In the future research, we plan to extend our study to the interval single-value neutrosophic sets

and do more dimensions practical applications.

Acknowledgments

This work was supported in part by the Natural Science Foundation of Fujian Province, China under Grant No. 2022J01958.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. L. A. Zadeh, Fuzzy sets, *Information and Control*, **8** (1965), 338–353. [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X)
2. M. W. Lin, H. B. Wang, Z. S. Xu, TODIM-based multi-criteria decision-making method with hesitant fuzzy linguistic term sets, *Artif. Intell. Rev.*, **53** (2020), 3647–3671. <https://doi.org/10.1007/s10462-019-09774-9>
3. M. W. Lin, C. Huang, Z. S. Xu, R. Q. Chen, Evaluation IoT platforms using integrated probabilistic linguistic MCDM method, *IEEE Internet Things J.*, **7** (2020), 11195–11208. <https://doi.org/10.1109/JIOT.2020.2997133>
4. M. W. Lin, X. M. Li, L. F. Chen, Linguistic q-rung orthopair fuzzy sets and their interactional partitioned Heronian mean aggregation operators, *Int. J. Intell. Syst.*, **35** (2020), 217–249. <https://doi.org/10.1002/int.22136>
5. M. W. Lin, C. Huang, R. Q. Chen, H. Fujita, X. Wang, Directional correlation coefficient measures for Pythagorean fuzzy sets: their applications to medical diagnosis and cluster analysis, *Complex Intell. Syst.*, **7** (2021), 1025–1043. <https://doi.org/10.1007/s40747-020-00261-1>
6. K. T. Atanassov, Intuitionistic fuzzy sets, *Fuzzy Sets and Systems*, **20** (1986), 87–96.
7. R. R. Yager, Generalized orthopair fuzzy sets, *IEEE Trans. Fuzzy Syst.*, **25** (2017), 1222–1230. <https://doi.org/10.1109/TFUZZ.2016.2604005>
8. W. Y. Zeng, H. S. Cui, Y. Q. Liu, Q. Yin, Z. S. Xu, Novel distance measure between intuitionistic fuzzy sets and its application in pattern recognition, *Iran. J. Fuzzy Syst.*, **19** (2022), 127–137. <https://doi.org/10.22111/IJFS.2022.6947>
9. H. B. Wang, F. Smarandache, R. Sunderraman, Single valued neutrosophic sets, *Multispace Multistruct.*, **4** (2010), 410–413.
10. N. Alpaslan, Neutrosophic set based local binary pattern for texture classification, *Expert Syst. Appl.*, **209** (2022), 118350. <https://doi.org/10.1016/j.eswa.2022.118350>
11. C. Huang, M. W. Lin, Z. S. Xu, Pythagorean fuzzy MULTIMOORA method based on distance measure and score function: its application in multicriteria decision making process, *Knowl. Inf. Syst.*, **62** (2020), 4373–4406. <https://doi.org/10.1007/s10115-020-01491-y>
12. M. W. Lin, X. M. Li, R. Q. Chen, H. Fujita, J. Lin, Picture fuzzy interactional partitioned Heronian mean aggregation operators: an application to MADM process, *Artif. Intell. Rev.*, **55** (2022), 1171–1208. <https://doi.org/10.1007/s10462-021-09953-7>
13. M. W. Lin, C. Huang, Z. S. Xu, MULTIMOORA based MCDM model for site selection of car sharing station under picture fuzzy environment, *Sustain. Cities Soc.*, **53** (2020), 101873. <https://doi.org/10.1016/j.scs.2019.101873>

14. M. Qiyas, M. Naeem, S. Khan, S. Abdullah, T. Botmart, T. J. M. Shah, Decision support system based on CoCoSo method with the picture fuzzy information, *J. Math.*, **2022** (2022), 1476233. <https://doi.org/10.1155/2022/1476233>
15. M. X. Luo, G. F. Zhang, L. X. Wu, A novel distance between single valued neutrosophic sets and its application in pattern recognition, *Soft Comput.*, **26** (2022), 11129–11137. <https://doi.org/10.1007/s00500-022-07407-y>
16. W. K. Hanna, N. M. Radwan, Heart disease patient risk classification based on neutrosophic sets, *Int. J. Business Intelligence and Data Mining*, **20** (2022), 93–106. <https://doi.org/10.1504/IJBIDM.2022.119961>
17. M. H. Hassan, S. M. Darwish, S. M. Elkaffas, An efficient deadlock handling model based on neutrosophic logic: case study on real time healthcare database systems, *IEEE Access*, **10** (2022), 76607–76621. <https://doi.org/10.1109/ACCESS.2022.3192414>
18. Y. Jin, M. Kamran, N. Salamat, S. Z. Zeng, R. H. Khan, Novel distance measures for single-valued neutrosophic fuzzy sets and their applications to multicriteria group decision-making problem, *J. Funct. Space.*, **2022** (2022), 7233420. <https://doi.org/10.1155/2022/7233420>
19. Y. Rong, W. Y. Niu, H. Garg, Y. Liu, L. Y. Yu, A hybrid group decision approach based on MARCOS and regret theory for pharmaceutical enterprise assessment under a single-valued neutrosophic scenario, *Systems*, **10** (2022), 106. <https://doi.org/10.3390/systems10040106>
20. Y. Q. Zeng, H. P. Ren, T. H. Yang, S. X. Xiao, N. Xiong, A novel similarity measure of single-valued neutrosophic sets based on modified Manhattan distance and its applications, *Electronics*, **11** (2022), 941. <https://doi.org/10.3390/electronics11060941>
21. M. W. Lin, Z. Chen, R. Chen, H. Fujita, Evaluation of startup companies using multicriteria decision making based on hesitant fuzzy linguistic information envelopment analysis models, *Int. J. Intell. Syst.*, **36** (2021), 2292–2322. <https://doi.org/10.1002/int.22379>
22. M. W. Lin, C. Huang, Z. S. Xu, TOPSIS method based on correlation coefficient and entropy measure for linguistic Pythagorean fuzzy sets and its application to multiple attribute decision making, *Complexity*, **2019** (2019), 6967390. <https://doi.org/10.1155/2019/6967390>
23. P. Rani, A. R. Mishra, Novel single-valued neutrosophic combined compromise solution approach for sustainable waste electrical and electronic equipment recycling partner selection, *IEEE Trans. Eng. Manage.*, **69** (2022), 3139–3153. <https://doi.org/10.1109/TEM.2020.3033121>
24. M. W. Lin, H. Wang, Z. S. Xu, Z. Yao, J. Huang, Clustering algorithms based on correlation coefficients for probabilistic linguistic term sets, *Int. J. Intell. Syst.*, **33** (2018), 2402–2424. <https://doi.org/10.1002/int.22040>
25. P. A. Ejegwa, S. Wen, Y. M. Feng, W. Zhang, N. Tang, Novel Pythagorean fuzzy correlation measures via Pythagorean fuzzy deviation, variance, and covariance with applications to pattern recognition and career placement, *IEEE Trans. Fuzzy Syst.*, **30** (2022), 1660–1668. <https://doi.org/10.1109/TFUZZ.2021.3063794>
26. Y. Jin, M. Hussain, K. Ullah, A. Hussain, A new correlation coefficient based on T-Spherical fuzzy information with its applications in medical diagnosis and pattern recognition, *Symmetry*, **14** (2022), 2317. <https://doi.org/10.3390/sym14112317>
27. Q. Zhang, J. P. Liu, J. H. Hu, Z. H. Yao, J. Yang, New correlation coefficients of Pythagorean fuzzy set and its application to extended TODIM method, *J. Intell. Fuzzy Syst.*, **43** (2022), 509–523. <https://doi.org/10.3233/JIFS-212323>
28. S. H. Gurmani, H. Y. Chen, Y. H. Bai, Extension of TOPSIS method under q-rung orthopair fuzzy hypersoft environment based on correlation coefficients and its applications to multi-attribute group decision-making, *Int. J. Fuzzy Syst.*, **25** (2023), 1–14. <https://doi.org/10.1007/s40815-022-01386-w>

29. M. Sajjad, W. Salabun, S. Faizi, M. Ismail, J. Watrobski, Statistical and analytical approach of multi-criteria group decision-making based on the correlation coefficient under intuitionistic 2-tuple fuzzy linguistic environment, *Expert Syst. Appl.*, **193** (2022), 116341. <https://doi.org/10.1016/j.eswa.2021.116341>
30. W. Zeng, R. Ma, Z. P. Liu, Y. Xi, Q. Yin, Z. S. Xu, Some novel distance measures between dual hesitant fuzzy sets and their application in medical diagnosis, *Int. J. Intell. Syst.*, **37** (2022), 8683–8671. <https://doi.org/10.1002/int.22960>
31. S. Singh, A. H. Ganie, Some novel q-rung orthopair fuzzy correlation coefficients based on the statistical viewpoint with their applications, *J. Ambient Intell. Human. Comput.*, **13** (2022), 2227–2252. <https://doi.org/10.1007/s12652-021-02983-7>
32. D. A. Chiang, N. P. Lin, Correlation of fuzzy sets, *Fuzzy Set. Syst.*, **102** (1999), 221–226. [https://doi.org/10.1016/S0165-0114\(97\)00127-9](https://doi.org/10.1016/S0165-0114(97)00127-9)
33. H. B. Mitchell, A correlation coefficient for intuitionistic fuzzy sets, *Int. J. Intell. Syst.*, **19** (2004), 483–490. <https://doi.org/10.1002/int.20004>
34. J. Ye, Multicriteria decision-making method using the correlation coefficient under single-valued neutrosophic environment, *Int. J. Gen. Syst.*, **42** (2013), 386–394. <https://doi.org/10.1080/03081079.2012.761609>
35. J. Ye, Improved correlation coefficients of single valued neutrosophic sets and interval neutrosophic sets for multiple attribute decision making, *J. Intell. Fuzzy Syst.*, **27** (2014), 2453–2462. <https://doi.org/10.3233/IFS-141215>
36. Q. C. Meng, N. S. Pang, S. Y. Zhao, J. W. Gao, Two-stage optimal site selection for waste-to-energy plant using single-valued neutrosophic sets and geographic information system based multi-criteria decision-making approach: a case study of Beijing, China, *Waste Manage.*, **156** (2023), 283–296. <https://doi.org/10.1016/j.wasman.2022.05.025>
37. S. S. Song, Z. H. Jia, J. Yang, N. K. Kasabov, A fast image segmentation algorithm based on saliency map and neutrosophic set theory, *IEEE Photonics J.*, **12** (2020), 1–16. <https://doi.org/10.1109/JPHOT.2020.3026973>
38. L. F. Gou, Y. Zhong, A new fault diagnosis method based on attributes weighted neutrosophic set, *IEEE Access*, **7** (2019), 117740–117748. <https://doi.org/10.1109/ACCESS.2019.2936494>
39. B. Huang, X. Yang, G. Feng, C. Guo, Relative measure-based approaches for ranking single-valued neutrosophic values and their applications, *Int. J. Mach. Learn. Cyber.*, **13** (2022), 1535–1552. <https://doi.org/10.1007/s13042-021-01464-9>
40. P. Majumdar, Neutrosophic sets and its applications to decision making, In: *Computational intelligence for big data analysis*, Cham: Springer, 2015, 97–115. https://doi.org/10.1007/978-3-319-16598-1_4
41. J. Ye, Vector similarity measures of simplified neutrosophic sets and their application in multicriteria decision making, *International Journal of Fuzzy Systems*, **16** (2014), 204–211.



AIMS Press

© 2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)