Contemporary Mathematics
https://ojs.wiserpub.com/index.php/CM/

UNIVERSAL WISER PUBLISHER

Research Article

Correlation Coefficient of Interval-Valued Fuzzy Sets with Interval-Valued Reference Functions and Applications in Medical Diagnosis and Selecting Appropriate Medicines

Fokrul Alom Mazarbhuiya[1*], M. Shenify[2], Janet Kez[1], A. S. Wungreiphi[1], Rupa Paul[1]

[1] School of Fundamental and Applied Sciences, Assam Don Bosco University, India
[2] College of Computer Science and IT, Albaha University, Albaha, 65799, Saudi Arabia
E-mail: fokrul.mazarbhuiya@dbuniversity.ac.in

Received: 21 November 2024; **Revised:** 7 February 2025; **Accepted:** 4 March 2025

Abstract: Fuzzy set theory (FST) has become a dominant model to deal with imprecision, uncertainty, vagueness, and ambiguity arising from many real-world applications. There are a couple of extensions and generalizations of FST, and an interval-valued FST is one of the important generalizations where the membership values are expressed in terms of closed intervals in [0, 1]. Finding the correlation coefficient of different FSTs can be an interesting research problem, as it has many real applications, from social science to medical science. This article first proposes a new extension of fuzzy set in terms of an interval-valued membership function and an interval-valued reference function and then proposes a new method of finding the correlation coefficient using the statistical parameters, like covariance and variance of the proposed fuzzy sets. The proposed formulae can be applied to both discrete and continuous sets of the universe of discourse. It has also been found that the correlation coefficient computed by the proposed formula lies in the closed interval of -1 and 1, which establishes that the correlation coefficient of the proposed fuzzy sets not only gives the strength of the relationship but also tells whether they are positively or negatively correlated. Furthermore, two practical applications along with numerical examples are discussed in detail. The result convincingly establishes the efficacy of the proposed formula.

Keywords: interval-valued fuzzy set, inter-valued reference function, membership interval, covariance of interval-valued fuzzy sets

MSC: 62A86

1. Introduction

Zadeh [1] first initiated the notion of linguistic variables in the field of mathematics. Eventually, the mathematics of fuzziness made inroads in almost all the directions of human knowledge. Accordingly, fuzzy was introduced in the field of statistics, not only redefining the statistical parameters in terms of fuzziness but also successfully applying in many branches of human knowledge [2–4]. Incorporating the idea from traditional statistics in [5–7], the correlation coefficient for intuitionistic fuzzy sets (IFS), a generalization of fuzzy sets, was defined, which not only demonstrates the relation of the IFSs but also the depth of their association. In [8], the authors proposed the idea of an interval-valued fuzzy set by

DOI: https://doi.org//10.37256/cm.6220256108

replacing the membership values of fuzzy sets with intervals lying in [0, 1]. The authors in [9] presented the idea of the correlation coefficient of interval-valued IFSs (IVIFS).

A new method for calculating the correlation coefficient of IVIFSs was given that was analogous to the cosine of the angle of intersection in a finite set and the probability space [10]. A detailed analysis of the interval-valued fuzzy numbers, their correlation coefficient, and energy information are recorded in [11]. The authors of [12] presented four formulae for calculating the correlation coefficient of interval-valued Pythagorean hesitant fuzzy sets (IVPHFSs) and Pythagorean hesitant fuzzy sets (PHFSs). They also demonstrated that the correlation between two IVPHFSs could be calculated using the comparison method of intervals and the least common multiple extension method. In [13], the authors proposed the definition of interval-valued linear Diophantine fuzzy weighted average and interval-valued linear Diophantine fuzzy weighted geometric aggregation operators, and based on these, a supplier selection-based multi-criteria decision-making was proposed. In [14], the authors proposed many features of the correlation and the weighted correlation coefficients of the IVIFSSs. In [15], the authors studied the problems related to group decision-making where the decision-makers supplied most of the information using the interval-valued intuitionistic fuzzy decision matrices. Using "expected interval", a method is proposed for the calculation of the correlation coefficient of the fuzzy numbers [16]. An efficient method of calculating the correlation coefficient of IFSs was discussed in [17].

In [18], the authors proposed a weighted correlation formula to determine the correlation coefficient of interval-valued Pythagorean fuzzy hypersoft sets and the proposed method was demonstrated by the real-life example of the corona virus disease (COVID)-19 pandemic. In [19], the authors proposed an interval quaternion number space-based method to extend the usefulness of complex IVIFSs, their order relations, and many interval quaternion number-based operations. A novel correlation coefficient under interval quaternion representation along with a core function was proposed in [19]. In [20], the authors proposed to study the correlation coefficient and the weighted correlation coefficient of the IVIFSSs along with their compulsory features. The correlation and weighted correlation coefficients for interval-valued Fermatean hesitant fuzzy sets were added to the four types of correlation coefficients of Fermatean hesitant fuzzy sets that the authors provided in [21]. To address the hesitancy, uncertainty, and the inconsistency in decision-making, the authors proposed two correlation coefficients of linguistic interval hesitant fuzzy sets with applications [22]. In [23], the authors proposed the correlation of dual interval-valued hesitant fuzzy sets. In [24], the authors used a central interval method to evaluate the correlation coefficient of fuzzy data. In [25], the authors proposed a Pearson correlation coefficient-based technique to evaluate decision-making and management plans using fuzzy interval measures. The authors of [26] have extended the traditional method to determine if a multivariate fuzzy dataset can be described with the help of a multivariate normal distribution.

In [27], the authors not only proposed a novel approach of finding the correlation coefficient of IFSs but also successfully applied the formula in multi-criteria decision making. In [28], the authors proposed a generalized and flexible method of finding the correlation coefficient of T-Spherical Fuzzy Sets and applied the method in medical diagnosis and investment decisions. In [29], a three-way approach of modified correlation coefficient of IFSs was given. Thao et al. [30] proposed to use a method on the hesitation margins in the process of computations to boost the output reliability. In [31], the authors proposed a nice formula of the correlation coefficient of spherical fuzzy sets and applied it successfully to project selection, pattern recognition, and medical diagnosis. The formula [31] gives the values in the interval of $[-1, 1]$. Authors in [32] proposed a statistical method for money investing schemes to aid in ranking the substitutes with correlation coefficient measures and Laplacian energy.

The authors of [33, 34] have proposed a novel definition of fuzzy numbers in terms of a reference function and a membership function. The membership values are then determined by computing the difference between the two functions mentioned above. In [35], authors put forward a nice formula of the correlation coefficient of fuzzy sets. The formula [35] uses the covariance defined using the membership values computed by taking differences of the membership function and the reference function associated with the fuzzy sets. It was found that the method agrees with most of the existing methods available in the literature i.e., the values lie in the interval of 0 and 1.

The objective of this paper is threefold, firstly, an extension of fuzzy set is proposed [34, 35] by defining both the membership function and the reference function as interval-valued on [0, 1] such that their difference will be the membership of the fuzzy set. Secondly, a nice formula for finding the correlation coefficient of such interval-valued fuzzy sets with the interval-valued reference function has been proposed. The formula can be expressed for the fuzzy sets over discrete as well as continuous sets of the universe of discourse. Finally, two real-life applications, along with numerical examples related to medical diagnosis and effective medicine selection for patients, are discussed to validate the method's efficacy. It is to be noted that the value of the correlation coefficient lies in the closed interval of -1 and 1. Therefore, the correlation coefficient evaluated by the proposed method not only provides us the strength of the relationship of the interval-valued fuzzy sets with the interval valued reference function but also demonstrates that the aforesaid fuzzy sets can be positively or negatively correlated, which in turn establishes its superiority over some of the existing methods.

The format of this article is as follows. The problem definition is described in Section 2. Section 3 discusses the suggested approach. The method's suggested uses are covered in Section 4 along with numerical examples. In Section 5, we wrap up our article with a succinct conclusion and suggestions for further research.

2. Problem definition

Let X be the universe of discourse and $\mu_A(x)$ and $\nu_A(x)$ be two functions on X such that $0 \leq \mu_A(x) \leq \nu_A(x) \leq 1$ and $0 \leq \mu_A(x) + \nu_A(x) \leq 1$, then a fuzzy number [33] can be redefined as follows

$$A = \{(x,\ \mu_A(x),\ \nu_A(x));\ x \in X\}$$

where $\mu_A(x) \in [0,\ 1]$ is the membership function and $\nu_A(x) \in [0,\ 1]$ is the reference function of A in X such that $\mu_A(x) - \nu_A(x)$ is the membership value for any $x \in X$. Any fuzzy set $A = \{x,\ \mu_A(x) \in\ [0,\ 1],\ x \in X\}$ [36] in the usual sense can be represented as

$$A = \{(x,\ \mu_A(x),\ 0),\ x \in X\}$$

Accordingly, the authors [33, 34] redefined most of the algebraic operations on fuzzy numbers using both membership and reference functions. However, almost all the works were done on the fuzzy numbers. In [35], the authors generalized the notion of [33, 34] to the general fuzzy set by defining it over both the continuous and the discrete sets of the universe of discourse. In discrete case, let $X = \{x_1,\ x_2,\ \ldots,\ x_\mathrm{n}\}$ be the discrete universe of discourse, then the fuzzy set A [35] is characterized as follows:

$$A = \{(x_1,\ \mu_A(x_1),\ \nu_A\,(x_1)),\ (x_2,\ (\mu_A(x_2),\ \nu_A\,(x_2)),\ \ldots,\ (x_n,\ (\mu_A(x_n),\ \nu_A\,(x_n))\},\ x_i \in X$$

or in other words, we can also express it as

$$A = \{(x_i,\ \mu_A(x_i),\ \nu_A\,(x_i)),\ x_i \in X;\ i = 1,\ 2,\ \ldots n\}$$

where $\mu_A(x_i) - \nu_A\,(x_i)$ = the membership value of x_i in A. Obviously the membership values are lying in between 0 and 1.

In any statistical analysis, the correlation coefficient plays a vital role in the measure of the strength of the linear relationship between two variables. Let $(a_1,\ b_1),\ (a_2,\ b_2),\ \ldots\ (a_\mathrm{n},\ b_\mathrm{n})$ be a sample of n-pairs of independent observations, then the sample correlation coefficient ρ_{ab} between a and b is given by [37]

$$\rho_{ab} = \frac{\sum_{i=1}^{n}(a_i - \overline{a})(b_i - \overline{b})}{\sqrt{\sum_{i=1}^{n}(a_i - \overline{a})^2 \sum_{i=1}^{n}(b_i - \overline{b})^2}} \tag{1}$$

where $\overline{a} = \frac{\sum_{i=1}^{n} a_i}{n}$ and $\overline{b} = \frac{\sum_{i=1}^{n} b_i}{n}$ are the sample means of a and b respectively. It has been found that ρ_{ab} always lies in $[-1, 1]$. A positive value indicates a tendency that a and b increase in the same direction, and a negative value indicates that in the opposite direction. In classical statistics, the observations must follow certain probability distributions. However, in real-life situations, the observations are sometimes expressed in a more convenient way using linguistic terms or, more precisely, fuzzy variables. In such a situation, measuring the correlation coefficient of two variables involving fuzziness is a challenging task.

Chiang and Lin [38] proposed the definition of fuzzy correlation coefficient as follows: Let $\{x_1, x_2, \ldots, x_n\}$ be a random sample from the universal set X, then the correlation coefficient bet the fuzzy sets $A = \{(x_j, \mu_A(x_j)), (x_j) \in X\}$ and $B = \{(x_j, \mu_B(x_j)), x_j \in X\}$ is given by [38]

$$\rho_{AB} = \frac{\frac{\sum_{j=1}^{n}(\mu_A(x_j) - \overline{\mu_A})(\mu_B(x_j) - \overline{\mu_B})}{n-1}}{\sqrt{\frac{\sum_{j=1}^{n}(\mu_A(x_j) - \overline{\mu_A})^2}{n-1} \frac{\sum_{i=1}^{n}(\mu_B(x_j) - \overline{\mu_B})^2}{n-1}}} \tag{2}$$

where $\overline{\mu_A} = \frac{\sum_{j=1}^{n} \mu_A(x_j)}{n}$ and where $\overline{\mu_B} = \frac{\sum_{j=1}^{n} \mu_B(x_j)}{n}$ are the sample means of the membership functions of A and B respectively.

Gerstenkorn and Manko [7], defined the correlation coefficient of the intuitionistic fuzzy sets [5] $A = \{(x_j, \mu_A(x_j), \nu_A(x_j)); x_j \in X\}$ and $B = \{(x_j, \mu_B(x_j), \nu_B(x_j)); x_j \in X\}$ in a finite set $X = \{x_1, x_2, \ldots x_n\}$ as follows:

$$\rho_{GM} = \frac{C_{GM}(A, B)}{\sqrt{T(A) \cdot T(B)}} \tag{3}$$

where $C_{GM}(A, B) = \sum_{j=1}^{n}(\mu_A(x_j)\mu_B(x_j) + \nu_A(x_j)\nu_B(x_j))$, $T(A) = \sum_{j=1}^{n}(\mu_A(x_j)^2 + \nu_A(x_j)^2)$ and $T(B) = \sum_{j=1}^{n}(\mu_B(x_j)^2 + \nu_B(x_j)^2)$. The formulae (2) and (3) give the value of the correlation coefficient between 0 and 1.

In 2024, Wungreiphi and Mazarbhuiya [35] proposed the correlation coefficient of fuzzy sets A and B defined in terms of membership functions and reference functions on a universe of discourse X as follows:

$$\rho_{AB} = \frac{cov(A, B)}{\sqrt{cov(A, A) \cdot cov(B, B)}} \tag{4}$$

where for finite case (discrete universe of discourse X),

$$cov(A, B) = \sum_{i=0}^{n}(\mu_A(x_i) - \nu_A(x_i))(\mu_B(x_i) - \nu_B(x_i)),\ cov(A, A) = \sum_{i=0}^{n}(\mu_A(x_i) - \nu_A(x_i))^2,$$

and

$$cov(B,\ B)=\sum_{i=0}^{n}\left(\mu_B\left(x_i\right)-\nu_B\left(x_i\right)\right)^2.$$

And for X $\subseteq$ R (real-line), $cov(A,\ B)=\int_X\left(\mu_A(x)-\nu_A(x)\right)\left(\mu_B(x)-\nu_B(x)\right)\mathrm{d}x,\ cov(A,\ A)=\int_X\left(\mu_A(x)-\nu_A(x)\right)^2\mathrm{d}x,$ and $cov(B,\ B)=\int_X\left(\mu_A(x)-\nu_A(x)\right)^2\mathrm{d}x.$

The formulae (2)-(4) give the value of the correlation coefficient between 0 and 1. However, later one computes the correlation coefficient of fuzzy sets with membership functions and reference functions, and it is applicable for both discrete as well as continuous sets of universe discourse.

In this article, we propose to generalize the fuzzy set given in [33, 34] by extending the membership function $\mu_A:X\rightarrow[0,\ 1]$ to an interval-valued function $\mu_A:X\rightarrow I([0,\ 1])$ and the reference function $\nu_A:X\rightarrow[0,\ 1]$ to $\nu_A:X\rightarrow I([0,\ 1])$ where $I([0,\ 1])$ denotes the collection of all closed intervals in [0, 1]. Using the above concept, an interval-valued fuzzy set with an interval-valued reference function (IVFSIVRF) can be redefined as follows

$$A=\left\{\left(x_{\mathrm{i}},\ \left[\mu_A^l\left(x_i\right),\ \mu_A^u(x_i)\right],\ \left[\nu_A^l\left(x_i\right),\ \nu_A^u(x_i)\right]\right),\ x_{\mathrm{i}}\in X\right\}$$

where, $0\leq\nu_A^l\left(x_i\right)\leq\mu_A^l\left(x_i\right)\leq1,\ 0\leq\nu_A^u\left(x_i\right)\leq\mu_A^u\left(x_i\right)\leq1,\ \mu_A^l\left(x_i\right)$ and $\mu_A^u(x_i)$ are respectively lower and upper bound of fuzzy interval-valued membership function, $\nu_A^l\left(x_i\right)$ and $\nu_A^u\left(x_i\right)$ are the lower and upper bound of the interval-valued reference function respectively.

To show the practical significance of the proposed fuzzy set, the following example of mineral bed [35] is used and is shown in Figure 1 below.

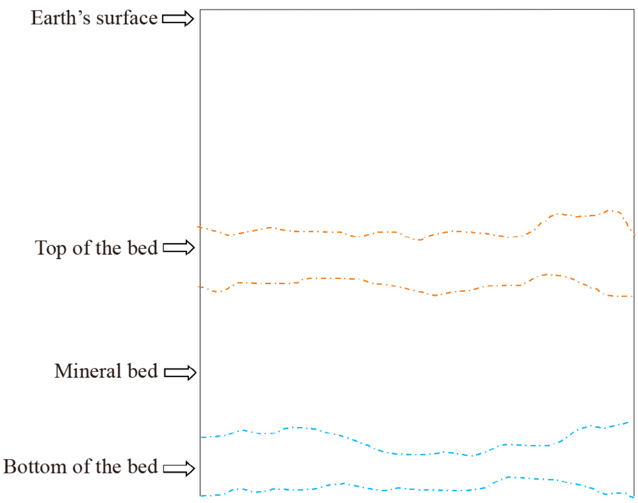

Figure 1. View of a mineral bed

The mineral deposit is found nearly 200-500 m below the surface of the ground, resulting in an almost 300-m-long mineral bed. Nearly, 200 m can be used to determine the depth of a mineral bed; not from the surface of the soil. Since, nearly, 200 is not a precise number, it would be convenient to represent as fuzzy number or more precisely fuzzy interval on [0, 1], therefore, every element in the surface will be lying in the interval of minimum and maximum membership value (reference value). Similar assumptions can be made for the membership function in the bottom of the bed. Letting 'x' be measured from a point on the surface of the earth, consider an inverted view of the mineral bed. The depth from the surface of the earth to the top of the bed is thus represented by $[d_1^l(x),\ d_1^u(x)]$, and the depth from the surface to the bottom of the bed is represented by $[d_2^l(x),\ d_2^u(x)]$. And the thickness of the mineral bed is represented by their difference.

3. Proposed method

The technique of computing the correlation coefficient of the proposed fuzzy sets over a discrete or a continuous universe of discourse is given as follows:

Let $A = \{(x_{\mathrm{i}},\ [\mu_A^l(x_i),\ \mu_A^u(x_i)],\ [\nu_A^l(x_i),\ \nu_A^u(x_i)]),\ x_{\mathrm{i}} \in X\}$, and $B = \{(x_{\mathrm{i}},\ [\mu_A^l(x_i),\ \mu_A^u(x_i)],\ [\nu_A^l(x_i),\ \nu_A^u(x_i)]),\ x_{\mathrm{i}} \in X\}$ be the two interval-valued fuzzy sets with IVFSIVRF over $X = \{x_1,\ x_2,\ \dots\ x_{\mathrm{n}}\}$ (discrete), then we define the correlation coefficient between A and B as follows:

$$\rho_{AB} = \frac{cov(A,\ B)}{\sqrt{cov(A,\ A) \cdot cov(B,\ B)}} \tag{5}$$

where $cov(A,\ B) = \sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))(\mu_A^u(x_i) - \nu_A^u(x_i))(\mu_B^l(x_i) - \nu_B^l(x_i))(\mu_B^u(x_i) - \nu_B^u(x_i))$ = covariance of A and B, $cov(A,\ A) = \sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))^2(\mu_A^u(x_i) - \nu_A^u(x_i))^2$ and $cov(B,\ B) = \sum_{i=0}^{n} (\mu_B^l(x_i) - \nu_B^l(x_i))^2(\mu_B^u(x_i) - \nu_B^u(x_i))^2$ are the covariance of A and the covariance of B, respectively.

Let $A = \{(x,\ [\mu_A^l(x), \mu_A^u(x)],\ [\nu_A^l(x),\ \nu_A^u(x)]),\ x \in X\}$, and $B = \{(x,\ [\mu_A^l(x),\ \mu_A^u(x)],\ [\nu_A^l(x), \nu_A^u(x)]),\ x \in X\}$ be the two IVFSIVRF over a universe of discourse $X \subseteq \mathbb{R}$, (real-line) then the correlation coefficient between A and B is given by the formula

$$\rho_{AB} = \frac{cov(A,\ B)}{\sqrt{cov(A,\ A) \cdot cov(B,\ B)}} \tag{6}$$

whereas $cov(A,\ B) = \int_X ((\mu_A^l(x) - \nu_A^l(x))(\mu_A^u(x) - \nu_A^u(x))(\mu_B^l(x) - \nu_B^l(x))(\mu_B^u(x) - \nu_B^u(x))\mathrm{d}x$ = covariance of A and B, $cov(A,\ A) = \int_X ((\mu_A^l(x) - \nu_A^l(x))(\mu_A^u(x) - \nu_A^u(x))^2\mathrm{d}x$, and $cov(B,\ B) = \int_X (\mu_B^l(x) - \nu_B^l(x))(\mu_B^u(x) - \nu_B^u(x))^2\mathrm{d}x$ are the covariance of A and the covariance of B, respectively.

Theorem 1 If $A = \{(x,\ [\mu_A^l(x),\ \mu_A^u(x)],\ [\nu_A^l(x),\ \nu_A^u(x)]),\ x \in X\}$, and $B = \{(x,\ [\mu_A^l(x),\ \mu_A^u(x)],\ [\nu_A^l(x),\ \nu_A^u(x)]),\ x \in X\}$ be the two IVFSIVRF over X (discrete or continuous), then the ρ_{AB} satisfies the following properties:

(i) $\rho_{AB} = 1$ if $A = B$.

(ii) $\rho_{AB} = \rho_{BA}$.

(iii) $-1 \leq \rho_{AB} \leq 1$.

Proof. (i) Let $A = \{(x_{\mathrm{i}},\ [\mu_A^l(x_i),\ \mu_A^u(x_i)],\ [\nu_A^l(x_i),\ \nu_A^u(x_i)]),\ x_{\mathrm{i}} \in X\}$, be an IVFSIVRF over $X = \{x_1,\ x_2,\ \dots\ x_{\mathrm{n}}\}$, then

$$\rho_{AA} = \frac{cov(A,\ A)}{\sqrt{cov(A,\ A) \cdot cov(A,\ A)}}$$

$$= \frac{\sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))(\mu_A^u(x_i) - \nu_A^u(x_i))(\mu_A^l(x_i) - \nu_A^l(x_i))(\mu_A^u(x_i) - \nu_A^u(x_i))}{\sqrt{\sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))^2(\mu_A^u(x_i) - \nu_A^u(x_i))^2.\ \sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))^2(\mu_A^u(x_i) - \nu_A^u(x_i))^2}}$$

$$\Rightarrow (\rho_{AA})^2 = \frac{\left(\sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))^2(\mu_A^u(x_i) - \nu_A^u(x_i))^2\right)^2}{\left(\sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))^2(\mu_A^u(x_i) - \nu_A^u(x_i))^2\right)^2} = 1$$

$$\Rightarrow \rho_{AA} = \pm 1$$

(ii) We have

$$\rho_{AB} = \frac{cov(A,\ B)}{\sqrt{cov(A,\ A) \cdot cov(B,\ B)}}$$

$$= \frac{\sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))(\mu_A^u(x_i) - \nu_A^u(x_i))(\mu_B^l(x_i) - \nu_B^l(x_i))(\mu_B^u(x_i) - \nu_B^u(x_i))}{\sqrt{\sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))^2(\mu_A^u(x_i) - \nu_A^u(x_i))^2 . \sum_{i=0}^{n} (\mu_B^l(x_i) - \nu_B^l(x_i))^2(\mu_B^u(x_i) - \nu_B^u(x_i))^2}}$$

$$= \frac{\sum_{i=0}^{n} (\mu_B^l(x_i) - \nu_B^l(x_i))(\mu_B^u(x_i) - \nu_B^u(x_i))(\mu_A^l(x_i) - \nu_A^l(x_i))(\mu_A^u(x_i) - \nu_A^u(x_i))}{\sqrt{\sum_{i=0}^{n} (\mu_B^l(x_i) - \nu_B^l(x_i))^2(\mu_B^u(x_i) - \nu_B^u(x_i))^2 \sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))^2(\mu_A^u(x_i) - \nu_A^u(x_i))^2.}}$$

$$= \rho_{BA}$$

In order to demonstrate the property (iii) let us take discrete case. Let $A = \{(x_{\mathrm{i}},\ [\mu_A^l(x_i),\ \mu_A^u(x_i)],\ [\nu_A^l(x_i),\ \nu_A^u(x_i)]),\ x_{\mathrm{i}} \in X\}$, and $B = \{(x_{\mathrm{i}},\ [\mu_A^l(x_i),\ \mu_A^u(x_i)], [\nu_A^l(x_i),\ \nu_A^u(x_i)]),\ x_{\mathrm{i}} \in X\}$ be the two IVFSIVRF over $X = \{x_1,\ x_2,\ \ldots\ x_{\mathrm{n}}\}$. Let us assume $\mu_B^l(x_{\mathrm{i}}) = a.\ \mu_A^l(x_{\mathrm{i}}) + b,\ \mu_B^u(x_{\mathrm{i}}) = a.\ \mu_A^u(x_{\mathrm{i}}) + b,\ \nu_B^l(x_{\mathrm{i}}) = a.\ \nu_A^l(x_{\mathrm{i}}) + b$, and $\nu_B^u(x_{\mathrm{i}}) = a.\ \nu_A^u(x_{\mathrm{i}}) + b$ for any two real numbers a and $b,\ i = 1,\ 2,\ \ldots\ n$. We choose a and b in such a way that the conditions $0 \leq \nu_A^l(x) \leq \mu_A^l(x) \leq 1$, $0 \leq \nu_A^u(x) \leq \mu_A^u(x) \leq 1$. Using (5), the correlation coefficient of A and B can be expressed as follows:

$$\rho_{AB} = \frac{cov(A,\, B)}{\sqrt{cov(A,\, A) \cdot cov(B,\, B)}}$$

$$= \frac{\sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))(\mu_A^u(x_i) - \nu_A^u(x_i))\left(a \cdot \mu_A^l(x_i) + b - a \cdot \nu_A^l(x_i) - b\right)(a \cdot\ \mu_A^u(x_i) + b - a \cdot \nu_A^u(x_i) - b)}{\sqrt{\sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))^2 (\mu_A^u(x_i) - \nu_A^u(x_i))^2 \cdot \sum_{i=0}^{n} (a \cdot \mu_A^l(x_i) + b - a \cdot \nu_A^l(x_i) - b)^2 (a \cdot \mu_A^u(x_i) + b - a \cdot \nu_A^u(x_i) - b)^2}}$$

$$= \frac{a^2 \sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))(\mu_A^u(x_i) - \nu_A^u(x_i))\left(\mu_A^l(x_i) - \nu_A^l(x_i)\right)(\mu_A^u(x_i) - \nu_A^u(x_i))}{\sqrt{a^2 \sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))^2 (\mu_A^u(x_i) - \nu_A^u(x_i))^2 \cdot \sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))^2 (\mu_A^u(x_i) - a \cdot \nu_A^u(x_i))^2}}$$

$$= \frac{a^2 \sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))^2 (\mu_A^u(x_i) - \nu_A^u(x_i))^2}{\sqrt{a^2 \cdot a^2 \left[\sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))^2 (\mu_A^u(x_i) - \nu_A^u(x_i))^2\right]^2}}$$

$$= \frac{a}{|a|}$$

$$= \left\{ \begin{array}{lll} 1 & \text{if} & a > 0 \\ -1 & \text{if} & a < 0 \end{array} \right\} \tag{7}$$

By Cauchy-schwarz inequality,

$$\left[\sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))(\mu_A^u(x_i) - \nu_A^u(x_i))(\ \mu_A^l(x_i) - \nu_A^l(x_i))(\mu_A^u(x_i) - \nu_A^u(x_i))\right]^2$$

$$\leq \sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))^2 (\mu_A^u(x_i) - \nu_A^u(x_i))^2 \sum_{i=0}^{n} (\ \mu_B^l(x_i) - \nu_B^l(x_i))^2 (\mu_B^u(x_i) - \nu_B^u(x_i))^2$$

$$\Longrightarrow \left| \frac{\sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))(\mu_A^u(x_i) - \nu_A^u(x_i))(\ \mu_A^l(x_i) - \nu_A^l(x_i))(\mu_A^u(x_i) - \nu_A^u(x_i))}{\left[\sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))^2 (\mu_A^u(x_i) - \nu_A^u(x_i))^2\right]^{\frac{1}{2}} \left[\sum_{i=0}^{n} (\ \mu_B^l(x_i) - \nu_B^l(x_i))^2 (\mu_B^u(x_i) - \nu_B^u(x_i))^2\right]^{\frac{1}{2}}} \right| \leq 1$$

$$\Longrightarrow \left| \frac{\sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))(\mu_A^u(x_i) - \nu_A^u(x_i))(\ \mu_A^l(x_i) - \nu_A^l(x_i))(\mu_A^u(x_i) - \nu_A^u(x_i))}{\sqrt{\sum_{i=0}^{n} (\mu_A^l(x_i) - \nu_A^l(x_i))^2 (\mu_A^u(x_i) - \nu_A^u(x_i))^2 \sum_{i=0}^{n} (\ \mu_B^l(x_i) - \nu_B^l(x_i))^2 (\mu_B^u(x_i) - \nu_B^u(x_i))^2}} \right| \leq 1$$

$$\Longrightarrow |\rho_{AB}| \leq 1 [\text{using } (5)]$$

Therefore, $-1 \leq \rho_{AB} \leq 1$. Hence proved. $\square$

4. Proposed applications

4.1 *Application in medical diagnosis*

To demonstrate the efficacy of our method, let us take the following example of medical diagnosis. Let $D = \{D_1, D_2, \ldots D_l\}$ be the set of diagnosis for the set of symptoms $S = \{S_1, S_2, \ldots S_m\}$ and let $P = \{P_1, P_2, \ldots P_n\}$ be the set of patients, then the algorithm for the diagnosis can be expressed as follows:

Step 1. Using the domain expert's knowledge, each symptom is represented using a pair (μ, ν), where μ and ν are the membership function and the reference function, respectively, and both are the interval-valued lying on [0, 1].

Step 2. For each patient $P_j (j = 1, 2, \ldots n)$, determine the correlation coefficient of the patients with the diagnosis i.e., $\rho_{D_i P_j} \in [-1, 1]$ using formula (5) where $D_i (i = 1, 2, \ldots l)$.

Step 3. For each $P_j (j = 1, 2, \ldots n)$, select the diagnosis D_i with maximum correlation coefficient value. Hence it can be stated that diagnosis of P_j is D_i.

Example 1 To determine the correct diagnosis D for a particular patient P having the set of symptoms S, the knowledge of a medical domain expert's is essential for expressing each element to represent with the help of an IVFSIVRF. Let us assume the set of symptoms S as follows:

$$S = \{S_1, S_2, S_3, S_4\} = \{Temperature, Headache, Stomachpain, Coldcough\}$$

The set of diagnosis D is as follows:

$$D = \{D_1, D_2, D_3\} = \{Viral fever, Malaria, Typhoid\}$$

And the set of patient P is as follows:

$$P = \{P_1, P_2, P_3\} = \{\textbf{\textit{Jenny}}, \textbf{\textit{Mike}}, \textbf{\textit{Thomas}}\}$$

Here, each symptom is described by membership function and reference function pair (μ, ν), where both μ and ν are intervals in [0, 1]. The characteristics of the symptoms in terms of diagnosis are expressed in the tabulated form in Table 1 below.

Table 1. Characteristics of symptoms for the considered diagnosis

	D_1	D_2	D_3
S_1	([0.05, 0.6], [0, 0.35])	([0.25, 0.8], [0.1, 0.55])	([0.12, 0.85], [0.1, 0.73])
S_2	([0.12, 0.92], [0.02, 0.46])	([0.11, 1], [0.1, 0.44])	([0.23, 0.87], [0.22, 0.34])
S_3	([0.24, 0.85], [0.12, 0.28])	([0.11, 0.99], [0.1, 0.87])	([0.33, 0.89], [0.21, 0.76])
S_4	([0.28, 0.75], [0.22, 0.36])	([0.45, 0.98], [0.24, 0.39])	([0.29, 0.69], [0.14, 0.36])

Similarly, the characteristics of the symptoms for the considered patients are given in the tabulated form in Table 2 below.

Table 2. Characteristics of symptoms for the considered patients

	S_1	S_2	S_3	S_4
P_1	([0.27, 0.87], [0.12, 0.31])	([0.32, 0.91], [0.12, 0.43])	([0.21, 0.89], [0.11, 0.69])	([0.43, 0.82], [0.21, 0.41])
P_2	([0.32, 0.82], [0.11, 0.33])	([0.21, 1], [0.12, 0.45])	([0.32,0.89], [0.21, 0.29])	([0.21, 0.97], [0.15, 0.24])
P_3	([0.21, 0.89], [0.11, 0.33])	([0.22, 0.95], [0.21, 0.55])	([0.32,0.85], [0.21, 0.66])	([0.22, 0.81], [0.21, 0.37])

The following are the IVFSIVRFs as computed from the Tables 1 and 2,

$$D_1 = \{(S_1,\ [0.05,\ 0.6],\ [0,\ 0.35]),\ (S_2,\ [0.12,\ 0.92],\ [0.02,\ 0.46]),\ (S_3,\ [0.24,\ 0.85],\ [0.12,\ 0.28]),$$

$$(S_4,\ [0.28,\ 0.75],\ [0.22,\ 0.36])\}$$

$$D_2 = \{(S_1,\ [0.25,\ 0.8],\ [0.1,\ 0.55]),\ (S_2,\ [0.11,\ 1],\ [0.1,\ 0.44]),\ (S_3,\ [0.11,\ 0.99],\ [0.1,\ 0.87]),$$

$$(S_4,\ ([0.45,\ 0.98],\ [0.24,\ 0.39])\}$$

$$D_3 = \{(S_1,\ [0.12,\ 0.85],\ [0.1,\ 0.73]),\ (S_2,\ [0.23,\ 0.87],\ [0.22,\ 0.34]),\ (S_3,\ [0.33,\ 0.89],\ [0.21,\ 0.76]),$$

$$(S_4,\ [0.29,\ 0.69],\ [0.14,\ 0.36])\}$$

$$P_1 = \{(S_1,\ [0.14,\ 0.85],\ [0.1,\ 0.82]),\ (S_2,\ [0.32,\ 0.82],\ [0.11,\ 0.33]),\ (S_3,\ [0.21,\ 0.89],\ [0.11,\ 0.33]),$$

$$(S_4,\ [0.33,\ 0.91],\ [0.21,\ 0.48])\}$$

$$P_2 = \{(S_1,\ [0.32,\ 0.91],\ [0.12,\ 0.43]),\ (S_2,\ [0.21,\ 1],\ [0.12,\ 0.45]),\ (S_3,\ [0.22,\ 0.95],\ [0.21,\ 0.55]),$$

$$(S_4,\ [0.21,\ 0.89],\ [0.20,\ 0.39])\}$$

$$P_3 = \{(S_1,\ [0.21,\ 0.89],\ [0.11,\ 0.69]),\ (S_2,\ [0.32,\ 0.89],\ [0.21,\ 0.29]),\ (S_3,\ [0.32,\ 0.85],\ [0.21,\ 0.66]),$$

$$(S_4,\ [0.33,\ 0.87],\ [0.21,\ 0.42])\}$$

Then, we apply the formula for correlation coefficients to derive diagnosis D_i of each patient P_j. The results are expressed in the tabular form in Table 3 below.

Table 3. Correlation coefficient of considered patients and considered diagnosis

	D_1	D_2	D_3
P_1	0.35653	0.42691	0.59471
P_2	0.41279	0.458427	0.142043
P_3	0.7668	0.67691	0.7223

From Table 3, the following observations can be made:

From the classical statistics, D_1 is weakly associated with P_1, D_2 is moderately associated with P_1, and D_3 is also moderately associated with P_1. Similarly, D_1 is moderately associated with P_2, D_2 is moderately associated with P_2, and D_3 is very weakly associated with P_2. Similarly, D_1, D_2, and D_3 have strong association P_1. Since the correlation coefficient $\rho_{D_3P_1}$ = 0.59471 is the highest in the row, the diagnosis of *Viral fever* is matching highly with the patient *Thomas*; therefore, among all the diagnoses *Viral fever* is the most appropriate for the patient *Thomas*. Also, $\rho_{D_2P_2}$ = 0.458427. Thus, D_2 is moderately associated with P_2; still, it is the highest in the second row; therefore, the diagnosis of *Malaria* can be a better match for the patient *Mike*. That is *Mike* has a high chance of having *Malaria*. Similarly, the diagnosis *Typhoid* is strongly matched with the patient *Thomas*, as the corresponding correlation coefficient value is the highest in that row. Similarly, *Viral fever* and *Malaria* are also strongly associated with *Thomas*, as the corresponding correlation coefficient values are also high. So *Thomas* has high chance of having any one of the following diseases: *Viral fever*, *Malaria*, or *Typhoid*.

4.2 *Application in appropriate medicine selection*

To demonstrate the efficacy of our method, let us take another example. Let $M = \{M_1, M_2, \ldots M_\mathrm{n}\}$ be the set of medicines for the set of symptoms $S = \{S_1, S_2, \ldots S_\mathrm{t}\}$ and let $P = \{P_1, P_2, \ldots P_\mathrm{m}\}$ be the set of patients, then the clinical effects of the medicines on the patients can be determined using the following steps:

Step 1. Using the domain expert's opinion, each symptom can be characterized by a pair (μ, ν), where μ and are the membership function and the reference function respectively, and both are interval-valued on [0, 1].

Step 2. For each patient $P_j (j = 1, 2, \ldots m)$, the correlation coefficient $\rho_{M_iP_j} \in [-1, 1]$ is calculated using (5) where $M_\mathrm{i} (i = 1, 2, \ldots n)$.

Step 3. For each $P_j (j = 1, 2, \ldots m)$, select the medicine M_e for which $\rho_{M_eP_j}$ is maximum that means M_e is maximally effective for P_j. Thus, it can be asserted that the medicine M_e is the most appropriate for the patient P_j.

Example 2 Let us take the following example to explain the above clinical effect of a medicine to a patient. To make a proper selection of an effective medicine M for a particular patient P with the given values of symptoms S, a medical knowledgebase is required which involves elements characterized in terms of interval-valued fuzzy sets with interval-valued reference functions. Let us assume that the set of symptoms S as follows:

$$S = \{S_1, S_2, S_3, S_4\} = \{Temperature, Headache, coldcough, Bodyache\}$$

The set of medicines M is as follows:

$$M = \{M_1, M_2, M_3, M_4\} = \{Tylenol, Aspirin, Cimetidine, Expectorant\}$$

And the set of patient P is as follows:

$$P = \{P_1,\ P_2,\ P_3,\ P_4\} = \{Jack,\ Jenny,\ Mike,\ Thomas\}$$

Here, each symptom is described by membership function and reference function pair $(\mu,\ \nu)$, where both μ and ν are intervals in [0, 1]. The characteristics of the symptoms with respect to medicines are expressed in the tabulated form in Table 4 below.

Table 4. Characteristics of symptoms for the considered medicines

	M_1	M_2	M_3	M_4
S_1	([0.1, 0.5], [0, 0.3])	([0.3, 0.7], [0.1, 0.5])	([0.2, 0.8], [0.1, 0.7])	([0.5, 0.9], [0, 0.7])
S_2	([0.3, 0.9], [0.2, 0.4])	([0.2, 1], [0.1, 0.4])	([0.3, 1], [0.2, 0.3])	([0, 1], [0, 0.1])
S_3	([0.2, 0.8], [0.1, 0.2])	([0.1, 0.9], [0, 0.8])	([0.3, 0.9], [0.1, 0.7])	([0.2, 0.9], [0.1, 0.2])
S_4	([0.3, 0.7], [0.2, 0.3])	([0.5, 0.9], [0.2, 0.3])	([0.2, 0.6], [0, 0.3])	([0.1, 0.9], [0, 0.8])

Similarly, the characteristics of the symptoms for the considered patients are given in the tabulated form in Table 5 below.

Table 5. Characteristics of symptoms for the considered medicines

	S_1	S_2	S_3	S_4
P_1	([0.2, 0.7], [0.1, 0.3])	([0.2, 0.9], [0.1, 0.4])	([0.1, 0.9], [0, 0.6])	([0.3, 0.8], [0.1, 0.4])
P_2	([0.2, 0.9], [0.1, 0.3])	([0.1, 1], [0, 0.4])	([0.2, 1], [0.1, 0.2])	([0.1, 1], [0, 0.2])
P_3	([0.1, 0.9], [0.1, 0.3])	([0.2, 0.9], [0.1, 0.5])	([0.2, 0.8], [0.1, 0.6])	([0.2, 0.8], [0.1, 0.3])
P_4	([0.3, 0.9], [0.1, 0.4])	([0.1, 0.9], [0, 0.3])	([0.3, 0.7], [0, 0.2])	([0.2, 0.9], [0.1, 0.2])

The following are the IVFSIVRFs as computed from the Tables 4 and 5,

$$M_1 = \{(S_1,\ [0.1,\ 0.5],\ [0,\ 0.3]),\ (S_2,\ [0.3,\ 0.9],\ [0.2,\ 0.4]),\ (S_3,\ [0.2,\ 0.8],\ [0.1,\ 0.2]),\ (S_4,\ [0.3,\ 0.7],\ [0.2,\ 0.3])\}$$

$$M_2 = \{(S_1,\ [0.3,\ 0.7],\ [0.1,\ 0.5]),\ (S_2,\ [0.2,\ 1],\ [0.1,\ 0.4]),\ (S_3,\ [0.1,\ 0.9],\ [0,\ 0.8]),\ (S_4,\ [0.5,\ 0.9],\ [0.2,\ 0.3])\}$$

$$M_3 = \{(S_1,\ [0.2,\ 0.8],\ [0.1,\ 0.7]),\ (S_2,\ [0.3,\ 1],\ [0.2,\ 0.3]),\ (S_3,\ [0.3,\ 0.9],\ [0.1,\ 0.7]),\ (S_4,\ [0.2,\ 0.6],\ [0,\ 0.3])\}$$

$$M_4 = \{(S_1,\ [0.5,\ 0.9],\ [0,\ 0.7]),\ (S_2,\ [0,\ 1],\ [0,\ 0.1]),\ (S_3,\ [0.2,\ 0.9],\ [0.1,\ 0.2]),\ (S_4,\ [0.1,\ 0.9],\ [0,\ 0.8])\}$$

$$P_1 = \{(S_1,\ [0.2,\ 0.7],\ [0.1,\ 0.3]),\ (S_2,\ [0.2,\ 0.9],\ [0.1,\ 0.4]),\ (S_3,\ [0.1,\ 0.9],\ [0,\ 0.6]),\ (S_4,\ [0.3,\ 0.8],\ [0.1,\ 0.4])\}$$

$$P_2 = \{(S_1,\ [0.2,\ 0.9],\ [0.1,\ 0.3]),\ (S_2,\ [0.1,\ 1],\ [0,\ 0.4]),\ (S_3,\ [0.1,\ 0.9],\ [0,\ 0.6]),\ (S_4,\ [0.3,\ 0.8],\ [0.1,\ 0.4])\}$$

$P_3 = \{(S_1,\ [0.1,\ 0.9],\ [0.1,\ 0.3]),\ (S_2,\ [0.2,\ 0.9],\ [0.1,\ 0.5]),\ (S_3,\ [0.2,\ 0.8],\ [0.1,\ 0.6]),\ (S_4,\ [0.2,\ 0.8],\ [0.1,\ 0.3])\}$

$P_4 = \{(S_1,\ [0.3,\ 0.9],\ [0.1,\ 0.4]),\ (S_2,\ [0.1,\ 0.9],\ [0,\ 0.3]),\ (S_3,\ [0.3,\ 0.7],\ [0,\ 0.2]),\ (S_4,\ [0.3,\ 0.7],\ [0,\ 0.2])\}$

Then, we apply the formula for finding the correlation coefficients to derive the efficacy of each medicine for each considered patients. The results are expressed in the tabular form in Table 6 below.

Table 6. Correlation coefficient of considered patients and considered medicines

	M_1	M_2	M_3	M_4
P_1	0.86374	0.930967	0.918084	0.527652
P_2	0.84891	0.885351	0.888055	0.603476
P_3	0.8618575	0.813000	0.889297	0.073131
P_4	0.91810	0.768047	0.8262329	0.7420416

From Table 6, the following observations can be made:

Tylenol, *Aspirin*, and *Cimetidine* are very strongly associated with the patient *Jack*, but the association between *Aspirin* and *Jack* is the strongest among all four medicines, and the *Expectorant* is moderately associated with the patient *Jack*. Also, the medicines *Tylenol*, *Aspirin*, and *Cimetidine* are very strongly associated with the patient *Jenny*, and the *Expectorant* is strongly associated with the patient *Jenny*, but the association between *Cimetidine* and *Jenny* is the strongest among all the medicines. Similarly, the medicines *Tylenol*, *Aspirin*, and *Cimetidine* are very strongly associated with the patient *Mike*, but the association between *Cimetidine* and *Mike* is the strongest among all the medicines, and the medicine *Expectorant* is very weakly associated with the patient *Mike*. Finally, the medicines *Tylenol*, and *Cimetidine* are very strongly associated with the patient *Thomas*, and *Aspirin*, and *Expectorant* have a strong association with the patient *Thomas*. Thus, among all the medicines, *Aspirin* is the most appropriate medicine for the patient *Jack*, and *Cimetidine* is the most appropriate among all the medicines for both *Jenny* and *Mike*. In other words, *Aspirin* is the most efficient medicine for the patient *Jack*, *Cimetidine* is the efficient medicine for the patients *Jenny* and *Mike*. Also, *Tylenol* is the most efficient medicine for the patient *Thomas*. Similarly, the medicine *Expectorant* is the least effective medicine for patient *Mike*; it is moderately effective for *Jack* and strongly effective for the patients *Jenny* and *Thomas*. This way, the effectiveness of any medicine with respect to a particular patient or a group of patients can be determined easily. Since we are considering only positive membership values in the example, we will only find the positive effect of a medicine, not the side effect.

In order to make the comparative analysis, we take the similar approach given [35], which uses a membership function and reference function to define a fuzzy set. In the first example of medical diagnosis, finding the correlation coefficient using the approach [35], the following order of diagnosis can be made. For the patient P_1, $D_3 > D_2 > D_1$; for the patient P_2, $D_2 > D_1 > D_3$; and for P_3, it is $D_1 > D_3 > D_2$. Thus the patient *Jenny* may be maximally affected by the disease *Typhoid* and minimally affected by the disease *Viral Fever*, *Mike* may be maximally affected by the disease *Malaria* and minimally affected by the *Typhoid*, and *Thomas* may be maximally affected by the *Viral Fever* and minimally by *Malaria*. The obtained results convincingly coincide with the results of the proposed approach. In the second example of selecting the appropriate medicines for the patients, the following order of preferences can be observed. For the patient P_1, the order of the preference is $M_2 > M_3 > M_1 > M_4$; for the patient P_2, the order is $M_3 > M_2 > M_1 > M_4$; for P_3, the order is $M_3 > M_1 > M_2 > M_4$; and for P_4, the order is $M_1 > M_3 > M_2 > M_4$. Thus, the medicine *Aspirin* is the most appropriate or effective for the patient *Jack*; *Cimetidine* is the most appropriate for both *Jenny* and *Mike*. However, *Tylenol* is the most effective for the patient *Thomas*. Thus, it has been found that the results obtained by the proposed approach coincide with the obtained results of one of the existing approaches [35], which validates the efficacy and the feasibility of the proposed method.

In this article, we have taken only the symptom set into consideration as input data in the universe of discourse while defining an interval-valued fuzzy set with an interval-valued reference function on the set of symptoms. It can also be extended to include other input features or attributes, medicine characteristics, patient profiles, and other important aspects. The proposed method of the correlation coefficient will work. In that case, each of the interval-valued fuzzy sets with interval-valued reference functions has to be suitably redefined, incorporating all the important input features associated with medicines and patients. However, there may be an issue with the weights of the attributes, i.e., all the attributes may not be equally weighted. In that situation, a weight function can be included in the formula that associates each input data with a weight value. The proposed correlation coefficient method will be slightly modified by multiplying the weight of each attribute with its corresponding input value. This will minimize the sensitivity of the input parameters on the proposed method.

5. Conclusions

In this article, a unique technique of finding the correlation coefficient of the IVFSIVRFs is proposed. Here, the definition of fuzzy set is extended by incorporating a reference function with its membership function, and the membership values are calculated by taking the difference of the aforesaid functions. In this article, the membership function and reference function are chosen to be interval-valued, lying on [0, 1]. We have named it interval-valued fuzzy sets with IVFSIVRF. Then, a correlation coefficient formula is put forth in terms of the covariance of the aforementioned fuzzy sets. The formula is applicable for both discrete and continuous sets of the universe of discourse. For the discrete set of universe of discourse, the summation is used, and for the continuous case, the integration is used. It has been found that the correlation coefficient value falls between -1 and 1, indicating both the degree of relationship between the proposed fuzzy sets and whether they are positively or negatively correlated. This indicates that our formula is more effective than some of the current approaches, where the correlation coefficient values fall between 0 and 1. Also, we have demonstrated our work with the help of real-life applications in medical diagnosis and effective medicine selection. Two numerical examples, one for each application, are given to justify the method's efficacy.

Although the proposed method for the correlation coefficient works nicely, it is not free from the limitations. First of all, some of the input parameters might dominantly influence the correlation coefficient values that make the method biased to the dominant attributes. Secondly, the formula does not take into consideration the non-membership aspect of fuzziness.

In the future, the following lines of work can be taken.

- In the future, the suitable modification of the method can be made to each input to improve the method's performance in terms of accuracy and reliability.
- In the future, the other types of statistical parameters are also studied to enhance the method's accuracy.

Conflict of interest

Authors declare that they have no conflict of interest.

References

[1] Zadeh LA. Fuzzy sets as basis of theory of possibility. *Fuzzy Sets and Systems*. 1978; 1(1): 3-28.
[2] Aczel J, Ptanzagl J. Remarks on the measurement of subjective probability and information. *Metrica*. 1967; 11(1): 91-105.
[3] Kandel A, Byatt WJ. Fuzzy sets, fuzzy algebra and fuzzy statistics. *Proceedings of the IEEE*. 1978; 66(12): 1619-1639.
[4] Teran P. Law of large numbers for possibilistic mean value. *Fuzzy Sets and Systems*. 2014; 245: 116-124.

[5] Atanassov KT. Intuitionistic fuzzy sets. *Fuzzy Sets and Systems*. 1986; 20(1): 87-96.
[6] Son MJ. Correlation of intuitionistic fuzzy sets. *The Journal of Korean Institute of Intelligent Systems*. 2007; 17(4): 546-549.
[7] Gerstenkorn T, Manko J. Correlation of intuitionistic fuzzy sets. *Fuzzy Sets and Systems*. 1991; 44(1): 29-43.
[8] Zadeh LA. The concept of a linguistic variable and its applications to approximate reasoning-I. *Information Sciences*. 1975; 8(3): 199-249.
[9] Bustince H, Burilo P. Correlation of interval-valued intuitionistic fuzzy sets. *Fuzzy Sets and Systems*. 1995; 74(2): 237-244.
[10] Zeng W, Wang J. Correlation coefficient of interval-valued intuitionistic fuzzy sets. In: *2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*. IEEE; 2011. p.98-102.
[11] Wang G, Li X. Correlation and information energy of interval valued fuzzy numbers. *Fuzzy Sets and Systems*. 1999; 103(1): 169-175.
[12] Zhang T, Zhang M, Li L, Wu Q, Zhou L. Correlation coefficients of interval-valued Pythagorean hesitant fuzzy sets and their applications. *IEEE Access*. 2020; 8: 9271-9286.
[13] Petchimuthu S, Riaz M, Kamacı H. Correlation coefficient measures and aggregation operators on interval-valued linear Diophantine fuzzy sets and their applications. *Computational and Applied Mathematics*. 2022; 41(8): 409.
[14] Zulqarnain RM, Xin XL, Saqlain M, Khan WA. TOPSIS method based on the correlation coefficient of interval-valued intuitionistic fuzzy soft sets and aggregation operators with their application in decision making. *Journal of Mathematics*. 2021; 2021(1): 6656858.
[15] Park DG, Kwun YC, Park-II JH, Park Y. Correlation coefficient of interval-valued intuitionistic fuzzy sets and its application to multiple attribute group decision making problem. *Mathematical and Computer Modelling*. 2009; 50(9-10): 1279-1293.
[16] Hung WL, Wu JW. A note on the correlation of fuzzy numbers by expected interval. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. 2001; 9(4): 517-523.
[17] Huang HL, Guo Y. An improved correlation coefficient of intuitionistic fuzzy sets. *Journal of Intelligent Systems*. 2019; 28(2): 231-243.
[18] Zulqarnain RM, Ma WX, Siddique I, Ahmad H, Askar S. A fair bed allocation during COVID-19 pandemic using TOPSIS technique based on correlation coefficient for interval-valued Pythagorean fuzzy hypersoft set. *Scientific Reports*. 2024; 14(1): 7678.
[19] Su Y, Gong Z, Qin N. Complex interval-value intuitionistic fuzzy sets: quaternion number representation, correlation coefficient and applications. *AIMS Mathematics*. 2024; 9(8): 19943-19966.
[20] Zulqarnain RM, Ma WX, Siddique I, Alburaikan A, Khalifa HA, Alanzi AM. Prioritization of thermal energy storage techniques using TOPSIS method based on correlation coefficient for interval-valued intuitionistic fuzzy hypersoft set. *Symmetry*. 2023; 15(3): 615.
[21] Demir I. Novel correlation coefficients for interval-valued Fermatean hesitant fuzzy sets with pattern recognition application. *Turkish Journal of Mathematics*. 2023; 47(1): 213-233.
[22] Xu YW, Meng FY, Wang N. Correlation coefficients of linguistic interval hesitant fuzzy sets and their application. *Iranian Journal of Fuzzy Systems*. 2019; 16(4): 65-81.
[23] Farhadinia B. Correlation for dual hesitant fuzzy sets and dual interval-valued hesitant fuzzy sets. *International Journal of Intelligent Systems*. 2014; 29(2): 184-205.
[24] Saneifard R, Saneifard R. Correlation coefficient between fuzzy numbers based on central interval. *Journal of Fuzzy Set Valued Analysis*. 2012; 2012: 1-9.
[25] Cheng YT, Yang CC. The application of fuzzy correlation coefficient with fuzzy interval data. *International Journal of Innovative Management, Information & Production*. 2013; 4(1): 65-71.
[26] Hesamiana G, Ghasem Akbarib M. Testing hypotheses for multivariate normal distribution with fuzzy random variables. *International Journal of Systems Science*. 2022; 53(1): 14-24.
[27] Bajaj J, Kumar S. A new intuitionistic fuzzy correlation coefficient approach with applications in multi-criteria decision making. *Decision Analytic Journal*. 2023; 9: 100340.
[28] Jin Y, Hussain M, Ullah K, Hussain A. A new correlation coefficient based on T-spherical fuzzy information with its applications in medical diagnosis and pattern recognition. *Symmetry*. 2022; 14(11): 2317.

[29] Ejegwa PA, Onyeke IC. Medical diagnostic analysis on selected patients based on modified Thao et al's correlation coefficient of intuitionistic fuzzy sets via an algorithmic approach. *Journal of Fuzzy Extension and Applications*. 2020; 1(2): 122-132.

[30] Thao NX, Ali M, Smarandache F. An intuitionistic fuzzy clustering algorithm based on a new correlation coefficient with application in medical diagnosis. *Journal of Intelligent & Fuzzy Systems*. 2019; 36(1): 189-198.

[31] Ali M, Ali W, Hussain I, Shah R. A novel correlation coefficient for spherical fuzzy sets and application in pattern recognition, medical diagnosis and mega project selection. *Heliyon*. 2025; 2025(1): 9164932.

[32] Kumar AN, Shaik SB. Correlation coefficient measure of intuitionistic fuzzy graphs with application in money investing schemes. *Computing and Informatics*. 2023; 42(2): 436-456.

[33] Baruah HK. Towards forming a field of fuzzy sets. *International Journal of Energy, Information and Communications*. 2011; 2(1): 16-20.

[34] Baruah HK. The theory of fuzzy sets: beliefs and realities. *International Journal of Energy, Information and Communications*. 2011; 2(2): 1-22.

[35] Wungreiphi AS, Mazarbhuiya FA. Correlation of fuzzy set using membership function and reference function. *ICIC Express Letters*. 2024; 18(11): 1221-1228.

[36] Klir GJ, Yuan B. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Upper Saddle River, NJ: Prentice Hall; 1995.

[37] Snedecor GW, Cochran WG. *Statistical Methods*. 6th ed. Iowa State University Press; 1967.

[38] Chiang DA, Lin NP. Correlation of fuzzy sets. *Fuzzy Sets and Systems*. 1999; 102(2): 221-226.