Applied Mathematics and Computational Intelligence
Volume 12, No 4, Nov 2023 [52 – 65]

Utilizing the Clustering Techniques using Distance-Based Similarity Measures of SVNS in Medical Diagnosis

Fatin Fadhlina Ahmad Riza[1], Norzieha Mustapha[2*], Nor Athirah Mansor[3], Nur Akmal Shafira Mazlan[4], Suriana Alias[5], Roliza Md Yasin[6]

[1,2,3,4,5,6]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Cawangan Kelantan, 18500, Machang, Kelantan

* Corresponding author: norzieha864@uitm.edu.my

Received: 2 Aug 2023
Accepted: 5 Oct 2023

ABSTRACT

The clustering techniques, combined with distance-based similarity measures of Single Valued Neutrosophic Sets (SVNS) are studied and applied in medical diagnosis. The study starts with reviewing SVNS' theoretical foundations, emphasising its ability to capture and handle ambiguous data. This study focuses on integrating distance-based similarity measurements to improve the clustering process, which has seen limited implementation thus far. The set of data includes three patients with five symptoms and three diagnoses. To deal with the data in medical diagnosis, each patient is diagnosed with a disease based on distance-based similarity measures. The disease with the highest similarity measure value indicates the recognized disease for that patient. Then, the diseases are clustered into different categories depend on the values of confidence level. The obtained results show that the suggested method enhances the precision of medical diagnosis significantly, especially in cases with ambiguity and uncertainty.

Keywords: Clustering algorithm, distance based similarity measure, medical diagnosis, neutrosophic set

1 INTRODUCTION

Due to the complexity of diverse diseases, doctors can learn a lot about medical diagnosis issues from modern medical technologies, but this information is sometimes imprecise and incomplete. Zadeh [1] initially introduced the concept of the Fuzzy Set (FS) as a solution to processing challenges. This concept represents each element's membership degree with a single number ranging from 0 to 1. Fuzzy Set theory is extensively employed in a numerous of fields, including engineering, economics, and medical diagnosis [2].

Atanassov [3] developed FS into intuitionistic fuzzy set (IFS) and Smarandache [4] proposed neutrosophics set (NS). NS can manage data that is partial, indeterminate, and inconsistent, while independently demonstrating the degree of truth-membership, the degree of indeterminacy-membership, and the degree of false-membership [5]. Recently, Alias and Mohamad [6] stated that neutrosophic theory has been proposed as a better alternative when FS and IFS cannot handle information indeterminacy, whereby there also exist the uncertain opinion decision makers such as voting condition, resulted vote, blank vote and against.

Broumi and Smarandache [7], Wang et al. [8], Das et al. [9] and Yang et al. [10] presented the notion of single-valued neutrosophic set (SVNS). Ye [11] recently demonstrated how the correlation coefficient and cross-entropy measure of SVNSs can be applied to decision-making problems. Thus, SVNS has proven to be a useful mathematical tool for tackling a range of real-world problems involving imprecise, ambiguous, and inconsistent data.

Clustering is a critical step in data mining, pattern recognition, machine learning, and microbiology research. The clustering techniques have proven their merit in managing high-dimensional data by identifying inherent patterns and grouping similar instances. These unsupervised learning algorithms are particularly beneficial when dealing with medical datasets that contain a wealth of information, but whose data points may not be clearly labeled or understood [12]. Clustering can help identify patterns that may signify the presence of a disease. For instance, the algorithm might group together patients with similar symptoms, which could lead to the discovery of a new disease. Additionally, if patients within a cluster show a trend of developing a specific disease, it could enable early prediction and prevention.

The clustering methods have been studied by researchers using a variety of tools throughout the past few decades. Clustering data information can be done using a variety of ways, including numerical information, interval-valued information, linguistic information, and so on. In fuzzy data analysis, fuzzy clustering analysis is a basic yet crucial tool. Ruspini [13] introduced the idea of fuzzy division and a fuzzy clustering strategy such as the Fuzzy C-means (FCM) clustering method and incorporates the fuzzy idea into hard clustering.

The application of clustering techniques in conjunction with SVNS, particularly those leveraging distance-based similarity measures, remains relatively less explored area. Our work aims to bridge this gap and illustrate the potential of such combination in enhancing the efficacy of medical diagnoses. Till now, only few studies focused on clustering techniques with SVNS. Ye [11] and Huang [14] introduced single-valued neutrosophic clustering approaches based on two distance-based SVNS similarity measures, and presented a clustering algorithm for grouping single-valued neutrosophic data using SVNS similarity measures.

The notion of similarity is crucial across virtually all scientific disciplines. Numerous approaches have been formulated to handle the metrics of similarity in neutrosophic sets. Majumdar and Samanta [15] proposed an entropy measure after introducing various similarity measures for SVNS based on distances, a matching function, and membership grades. Ye [11] addressed and demonstrated the application of Hamming and Euclidean distance-based similarity measures on interval neutrosophic sets in problem-oriented decision-making scenarios. Pramanik et al. [16] also examined and analysed the application of hybrid vector similarity measures to interval neutrosophic sets, offering a thorough grasp of similarity measures in the domain of neutrosophic sets.

As stated by Shahzadi et al. [17], issues with medical diagnosis, signs and symptoms of various diseases may evolve over time. Many researchers proposed theories regarding neutrosophic sets in medical diagnosis. For instance, Mustapha et al. [18] proposed a similarity measure based on distance for use in medical diagnosis. In addition, Ye et al. [19] introduced improved cosine similarity measures for both Single-Valued Neutrosophic Sets (SVNSs) and interval neutrosophic sets in the field of medical diagnosis. Mondal and Pramanik [20] also proposed the tangent similarity measure and the weighted

tangent similarity measure for SVNSs, with applications in medical diagnosis.

A variety of methods and algorithms for addressing the medical diagnosis problem in a neutrosophic environment have been developed (Ye et al. [19]; Luo and Zhao [21]; Xiao [22], Mustapha et al. [18, 23] and Chai et al. [2]). Distance-based similarity metrics of single-valued neutrosophic multisets were presented for medical diagnosis Ye et al. [19]. Following that, Luo and Zhao [21] introduced a measure of distance between IFS and issues in medical diagnosis, whereas Xiao [22] explored the use of hybrid fuzzy sets within the context of medical diagnosis. Chai et al. [2] also proposed novel similarity measures for Single-Valued Neutrosophic Sets (SVNSs), with applications in medical diagnosis and pattern recognition.

This paper discusses the utilization of similarity measure of SVNS and clustering techniques in healthcare. We outline the underlying principles of SVNS, elucidate our proposed method of integrating these principles with distance-based similarity measures in clustering, and present our findings based on extensive data analyses in a medical context. This study aims to utilize the clustering algorithm following Ye [11] using distance-based similarity measures of SVNS and apply to medical diagnosis. Two similarity measures are chosen from Ye [11] and another one similarity measure is chosen from Mustapha et al. [23] for comparative study.

2 PRELIMINARIES

In this section, some of the preliminary concepts that must be understood in order to fully benefit from this study are as follows.

2.1 Single Valued Neutrosophic Set

SVNS is a neutrosophic set that can be used in real scientific and engineering applications.

Definition 1 [7]. Consider X as a set of points (objects), where x is a generic element in X. A truth membership function defines SVNS A in X, $T_A(x)$, an indeterminacy membership function, $I_A(x)$, and a falsity membership function $F_A(x)$. Here $[T_A(x), I_A(x), F_A(x)]$ are real subsets of $[0, 1]$. Thus, an SVNS A can be denoted by

$$A = \{\langle x, T_A(x), I_A(x), F_A(x)\rangle | x \in X\}. \tag{1}$$

2.2 Distance Based Similarity Measure of SVNS

A critical operation within the realm of SVNS is the measurement of similarity, which allows for comparing and differentiating between SVNSs. Distance-based similarity measures form one such important category of similarity measures in SVNS. Distance-based similarity measures which are using in this study are as follows:

Definition 2 [11] Consider a universal set $X = \{x_1, x_2, ..., x_n\}$ represented as such $A = \{< x_i, T_A(x_i), I_A(x_i), F_A(x_i) > |x_i \in X\}$ and $B = \{< x_i, T_B(x_i), I_B(x_i), F_B(x_i) > |x_i \in X\}$, where $T_A(x_i), I_A(x_i), F_A(x_i), T_B(x_i), I_B(x_i), F_B(x_i) \in [0, 1]$ for every $x_i \in X$. Let the weight $w_i (i = 1, 2, ..., n)$ of an element $x_i (i = 1, 2, ..., n)$, where $w_i \geq 0 (i = 1, 2, ..., n)$ and $\sum_{i=1}^{n} w_i = 1$.The single valued neutrosophic similarity measures are there-

fore defined as follows:

1) [11] $S_1(A,B) = 1 - d_p(A,B)$ (2)

2) [11] $S_2(A,B) = \dfrac{1 - d_p(A,B)}{1 + d_p(A,B)}$ (3)

3) [23] $S_3(A,B) = 1 - d_q(A,B)$ (4)

where

$$d_p(A,B) = \left\{ \frac{1}{3} \sum_{i=1}^{n} w_i \Big[|T_A(x_i) - T_B(x_i)| + |I_A(x_i) - I_B(x_i)| + |F_A(x_i) - F_B(x_i)| \Big] \right\} \tag{5}$$

and

$$d_q(A,B) = \frac{2}{n} \sum_{i=1}^{n} w_i \frac{sin\{\frac{\pi}{10}|T_A(x_i) - T_B(x_i)|\} + sin\{\frac{\pi}{10}|I_A(x_i) - I_B(x_i)|\} + sin\{\frac{\pi}{10}|F_A(x_i) - F_B(x_i)|\}}{1 + sin\{\frac{\pi}{10}|T_A(x_i) - T_B(x_i)|\} + sin\{\frac{\pi}{10}|I_A(x_i) - I_B(x_i)|\} + sin\{\frac{\pi}{10}|F_A(x_i) - F_B(x_i)|\}} \tag{6}$$

where w_i is a weight, $d_p(A,B)$ and $d_q(A,B)$ are the generalized single-valued neutrosophic weighted distance measures.

Proposition 1. Let A and B be two SVNSs in a discourse universe $X = \{x_1, x_2, \dots, x_n\}$; $S_n(A,B)$ is known as a single valued neutrosophic similarity measure, and it must meet the following requirements: $(C1) 0 \leq S_n(A,B) \leq 1$;
$(C2) S_n(A,B) = 1$ if and only if $A = B$;
$(C3) S_n(A,B) = S_n(B,A)$;
$(C4) S_n(A,C) \leq S_n(A,B)$ and $S_n(A,C) \leq S_n(B,C)$ if C is neutrosphic set in X and $A \subseteq B \subseteq C$.

The proved of $S_1(A,B)$ and $S_2(A,B)$ are discussed in [11] and $S_3(A,B)$ in Mustapha et al. [23].

3 METHODOLOGY

3.1 Clustering Algorithm using Distance-Based Similarity Measures of SVNSs

Such a clustering algorithm calculates the distance between data points represented as SVNSs using a distance-based similarity measure. The algorithm then groups data points into clusters based on these distance calculations. Then, the steps of clustering algorithm are given as follows:

1. Let $A = (A_1, A_2, ..., A_m)$ is a set of SVNSs, we may calculate the similarity measure degree of SVNSs using equations (2), (3) and (4). Then there's the similarity matrix $M = (s_{ij})_{m \times m}$, where $s_{ij} = S_k(A_i, A_j)(k = 1, 2)$ for $i, j = 1, 2, \dots, m$, where m represents cardinality of NS.

2. The process of constructing composition matrices is repeated until it is proven that

$$M \to M^2 \to M^4 \to \ldots \to M^{2^k} = M^{2^{(k+1)}},$$

which implies that M^{2^k} is a matrix of equivalent similarity, which $M^2 = M \circ M = (\bar{s}_{ij})_{m\times m} = max_k\{min(s_{ik}, s_{kj})\}_{m\times m}$.

3. For the equivalent similarity matrix $\overline{M} = (\bar{s}_{ij})_{m\times m}$, we can construct a λ-cutting matrix $\overline{M}_\lambda = (\overline{s}^{\lambda}_{ij})_{m\times m}$ of $\overline{M}$, where

$$\overline{s}^{\lambda}_{ij} = \begin{cases} 0, & \bar{s}_{ij} < \lambda; \\ 1, & \bar{s}_{ij} \geq \lambda, \end{cases} \text{ for } i,j = 1,2,\ldots,m, \tag{7}$$

and λ is the confidence level with $\lambda \in [0,1]$.

3.2 Practical Example

This practical example only showed for $S_3(A,B)$. Let the following three SVNSs in a universe of discourse $X = \{x_1, x_2\}$:

$$A = \{< x_1, 0.1, 0.5, 0.6 >, < x_2, 0.2, 0.5, 0.7 >\},$$
$$B = \{< x_1, 0.3, 0.4, 0.5 >, < x_2, 0.5, 0.3, 0.4 >\},$$
$$C = \{< x_1, 0.6, 0.1, 0.2 >, < x_2, 0.8, 0.1, 0.3 >\}$$

Then, by applying Equation (4), and the weight vector $w = (0.5, 0.5)^T$, the proposition 1 is satisfied.
(C1) $0 \leq S_n(A,B) \leq 1$;

$$\begin{aligned} S_3(A,B) &= 1 - d_q(A,B) \\ &= 1 - \Big(\frac{1}{2}\frac{sin\{\frac{\pi}{10}|0.1-0.3|\} + sin\{\frac{\pi}{10}|0.5-0.4|\} + sin\{\frac{\pi}{10}|0.6-0.5|\}}{1 + sin\{\frac{\pi}{10}|0.1-0.3|\} + sin\{\frac{\pi}{10}|0.5-0.4|\} + sin\{\frac{\pi}{10}|0.6-0.5|\}} \\ &+ \frac{1}{2}\frac{sin\{\frac{\pi}{10}|0.2-0.5|\} + sin\{\frac{\pi}{10}|0.5-0.3|\} + sin\{\frac{\pi}{10}|0.7-0.4|\}}{1 + sin\{\frac{\pi}{10}|0.2-0.5|\} + sin\{\frac{\pi}{10}|0.5-0.3|\} + sin\{\frac{\pi}{10}|0.7-0.4|\}}\Big) \\ &= 0.312 \end{aligned}$$

(C2) $S_n(A,B) = 1$ if and only if $A = B$;

If $A = B$,

$$\begin{aligned} S_3(A,A) &= 1 - d_q(A,A) \\ &= 1 - \Big(\frac{1}{2}\frac{sin\{\frac{\pi}{10}|0.1-0.1|\} + sin\{\frac{\pi}{10}|0.5-0.5|\} + sin\{\frac{\pi}{10}|0.6-0.6|\}}{1 + sin\{\frac{\pi}{10}|0.1-0.1|\} + sin\{\frac{\pi}{10}|0.5-0.5|\} + sin\{\frac{\pi}{10}|0.6-0.6|\}} \\ &+ \frac{1}{2}\frac{sin\{\frac{\pi}{10}|0.2-0.2|\} + sin\{\frac{\pi}{10}|0.5-0.5|\} + sin\{\frac{\pi}{10}|0.7-0.7|\}}{1 + sin\{\frac{\pi}{10}|0.2-0.2|\} + sin\{\frac{\pi}{10}|0.5-0.5|\} + sin\{\frac{\pi}{10}|0.7-0.7|\}}\Big) \\ &= 1 - 0 = 1 \end{aligned}$$

Thus, $S_3(A,B) = S_3(A,A)$ if and only if $A = B$;

(C3) $S_n(A,B) = S_n(B,A)$;

$$\begin{aligned} S_3(A,B) &= 0.312 \\ S_3(B,A) &= 1 - d_1(B,A) \\ &= 1 - \Big(\frac{1}{2}\frac{sin\{\frac{\pi}{10}|0.3-0.1|\} + sin\{\frac{\pi}{10}|0.4-0.5|\} + sin\{\frac{\pi}{10}|0.5-0.6|\}}{1 + sin\{\frac{\pi}{10}|0.3-0.1|\} + sin\{\frac{\pi}{10}|0.4-0.5|\} + sin\{\frac{\pi}{10}|0.5-0.6|\}} \\ &+ \frac{sin\{\frac{\pi}{10}|0.5-0.2|\} + sin\{\frac{\pi}{10}|0.3-0.5|\} + sin\{\frac{\pi}{10}|0.4-0.7|\}}{1 + sin\{\frac{\pi}{10}|0.5-0.2|\} + sin\{\frac{\pi}{10}|0.3-0.5|\} + sin\{\frac{\pi}{10}|0.4-0.7|\}}\Big) \\ &= 0.312 \end{aligned}$$

Hence, $S_3(A,B) = S_3(B,A) = 0.312$.

(C4) $S_n(A,C) \leq S_n(A,B)$ and $S_n(A,C) \leq S_n(B,C)$ if C is neutrosphic set in X and $A \subseteq B \subseteq C$;

$$\begin{aligned} S_3(A,C) &= 1 - d_q(A,C) \\ &= 1 - \Big(\frac{1}{2}\frac{sin\{\frac{\pi}{10}|0.1-0.6|\} + sin\{\frac{\pi}{10}|0.5-0.1|\} + sin\{\frac{\pi}{10}|0.6-0.2|\}}{1 + sin\{\frac{\pi}{10}|0.1-0.6|\} + sin\{\frac{\pi}{10}|0.5-0.1|\} + sin\{\frac{\pi}{10}|0.6-0.2|\}} \\ &+ \frac{sin\{\frac{\pi}{10}|0.2-0.8|\} + sin\{\frac{\pi}{10}|0.5-0.1|\} + sin\{\frac{\pi}{10}|0.7-0.3|\}}{1 + sin\{\frac{\pi}{10}|0.2-0.8|\} + sin\{\frac{\pi}{10}|0.5-0.1|\} + sin\{\frac{\pi}{10}|0.7-0.3|\}}\Big) \\ &= 0.703 \end{aligned}$$

$$
\begin{aligned}
S_3(B,C) &= 1 - d_q(B,C) \\
&= 1 - \Big(\frac{1}{2}\frac{sin\{\frac{\pi}{10}|0.3-0.6|\} + sin\{\frac{\pi}{10}|0.4-0.1|\} + sin\{\frac{\pi}{10}|0.5-0.2|\}}{1 + sin\{\frac{\pi}{10}|0.3-0.6|\} + sin\{\frac{\pi}{10}|0.4-0.1|\} + sin\{\frac{\pi}{10}|0.5-0.2|\}} \\
&+ \frac{sin\{\frac{\pi}{10}|0.5-0.8|\} + sin\{\frac{\pi}{10}|0.3-0.1|\} + sin\{\frac{\pi}{10}|0.4-0.3|\}}{1 + sin\{\frac{\pi}{10}|0.3-0.1|\} + sin\{\frac{\pi}{10}|0.4-0.3|\} + sin\{\frac{\pi}{10}|0.3-0.1|\}}\Big) \\
&= 0.811
\end{aligned}
$$

Thus, $S_3(A,C) \leq S(A,B)$ and $S_3(A,C) \leq S_3(B,C)$.

3.3 Numerical example in medical diagnosis

Medical diagnosis data are collected from Shahzadi et al. [17] and shown in Tables 1 - 2. Consider Patient 1, Patient 2, and Patient 3 as three patients with certain symptoms. The patients' symptoms include temperature(T), insulin(I), blood pressure(BP), blood platelets(BPL) and cough(C). The diagnoses are diabetes(D), dengue fever(DF), and tuberculosis(TB). The data in Tables 1 and 2 demonstrates the relationship between patients and symptoms as well as the relationship between symptoms and diagnoses.

There are three specified degree of membership function for each time inspection which is for truth description of symptom, indeterminacy description symptom and falsity description symptom. As can be seen from the tables, the truth membership degree for T which belong to Patient 1 is equal to 0.8. The indeterminate membership degree for T which belong to Patient 1 is equal to 0.1 and the falsity membership degree for T which belong to patient 1 is equal to 0.1. Patient 1 has a high symptom T since the truth value is equal to 0.8. Meanwhile, Patient 3 with falsity value equal 0.7, indicate that Patient 3 possibility has no BP problem.

Suppose the standard relation between disease and symptoms as represented in Table 2, the truth membership degree for T which belong to D is equal to 0.2, the indeterminate membership degree for T which belong to D is equal to 0.0, and the falsity membership degree for T which belong to D is equal to 0.8. Referring to the relationship between D and the five symptoms, the highest truth membership value is equal to 0.9 which belongs to insulin problem indicating one has diabetes.

The data of medical diagnosis in Tables 1 - 2 are applied to $S_1(A,B)$, $S_2(A,B)$ and $S_3(A,B)$. By applying Equations (2) - (4), we can determine the similarity measures of SVNSs. Results from Shahzadi et al. [17] are used to validate the results to ensure they are practical and effective.

Table 1 : The relation between patients and symptoms

Relation	Patient 1	Patient 2	Patient 3
T	(0.8,0.1,0.1)	(0.6,0.2,0.2)	(0.4,0.2,0.4)
I	(0.2,0.2,0.6)	(0.9,0.0,0.1)	(0.2,0.1,0.7)
BP	(0.4,0.2,0.4)	(0.1,0.1,0.8)	(0.1,0.2,0.7)
BPL	(0.8,0.1,0.1)	(0.2,0.1,0.7)	(0.3,0.1,0.6)
C	(0.3,0.3,0.4)	(0.5,0.1,0.4)	(0.8,0.0,0.2)

Table 2 : The relation between symptoms and diagnoses

Relation	T	I	BP	BPL	C
D	(0.2,0.0,0.8)	(0.9,0.0,0.1)	(0.1,0.1,0.8)	(0.1,0.1,0.8)	(0.1,0.1,0.8)
DF	(0.9,0.0,0.1)	(0.0,0.2,0.8)	(0.8,0.1,0.1)	(0.9,0.0,0.1)	(0.1,0.1,0.8)
TB	(0.6,0.2,0.2)	(0.0,0.1,0.9)	(0.4,0.2,0.4)	(0.0,0.2,0.8)	(0.9,0.0,0.1)

4 RESULTS AND DISCUSSION

4.1 Solutions using distance based similarity measures

The similarity is important in recognizing the element set properties. Tables 3 - 5 show the results of similarity measures using $S_1(A,B)$, $S_2(A,B)$ and $S_3(A,B)$. A similarity measure value close to 1 indicates that the two NSs being compared are nearly identical in their truth-membership, indeterminacy-membership, and falsity-membership functions [24]. As diagnosis results in Table 3 - 5, we can see that for Patient 1, $S_3(A,B)$ has closer value to 1 which is 0.8663 compared to the values obtained by $S_1(A,B)$ and $S_2(A,B)$. This suggests that $S_3(A,B)$ is more representable for medical diagnosis problem.

The highest value of the similarity measure in each column determines best medical diagnosis. Table 6 depicts that Patient 1 suffers from dengue, Patient 2 suffers from diabetes, and Patient 3 suffers from tuberculosis using $S_1(A,B)$, $S_2(A,B)$ and $S_3(A,B)$. The results in Table 6 also similar with the finding from [17] and this implies that the suggested similarity measures are both viable and effective.

Table 3 : The similarity measure between patients and symptoms using S_{1SVNS}

Diagnosis	Patient 1	Patient 2	Patient 3
D	0.6133	**0.8533**	0.7200
DF	**0.8400**	0.6000	0.6400
TB	0.7467	0.7467	**0.8533**

Table 4 : The similarity measure between patients and symptoms using S_{2SVNS}

Diagnosis	Patient 1	Patient 2	Patient 3
D	0.4423	**0.7442**	0.5625
DF	**0.7241**	0.4286	0.4706
TB	0.5957	0.5957	**0.7442**

Table 5 : The similarity measure between patients and symptoms using S_{3SVNS}

Diagnosis	Patient 1	Patient 2	Patient 3
D	0.7828	**0.8733**	0.8146
DF	**0.8663**	0.7883	0.7986
TB	0.8301	0.8302	**0.8828**

Table 6 : Summary table

	Patient 1	Patient 2	Patient 3
S_{1SVNS}	DF	D	TB
S_{2SVNS}	DF	D	TB
S_{3SVNS}	DF	D	TB
Shahzadi et al. [17]	DF	D	TB

4.2 Clustering analysis

In this section, we will present and discuss the results of clustering algorithms using $S_1(A,B)$, $S_2(A,B)$ and $S_3(A,B)$.

4.2.1 Clustering analysis Using $S_1(A,B)$

The first step is to create the similarity matrix as shown below:

$$M = \begin{bmatrix} 1 & 0.7733 & 0.8666 \\ 0.7733 & 1 & 0.9067 \\ 0.8666 & 0.9067 & 1 \end{bmatrix}.$$

The elements in similarity matrix are calculated from Table 3, follow the steps in Section 3.1 as shown below:

$M_{11} = 1 - |d_p(Patient1, D) - d_p(Patient1, D)| = 1 - |0.3867 - 0.3867| = 1;$

$M_{12} = 1 - |d_p(Patient1, D) - d_p(Patient1, DF)| = 1 - |0.3867 - 0.1600| = 0.7733;$

$M_{13} = 1 - |d_p(Patient1, D) - d_p(Patient1, TB)| = 1 - |0.3867 - 0.2533| = 0.8666;$

$M_{23} = 1 - |d_p(Patient1, DF) - d_p(Patient1, TB)| = 1 - |0.6100 - 0.2533| = 0.9067;$

$M_{11} = M_{22} = M_{33}; M_{12} = M_{21}; M_{13} = M_{31}; M_{23} = M_{32}.$

The second step is to find the limited time compositions of identical similarity matrices M:

$$M^2 = \begin{bmatrix} 1 & 0.7733 & 0.8666 \\ 0.7733 & 1 & 0.9067 \\ 0.8666 & 0.9067 & 1 \end{bmatrix},$$

It is clearly that $M^2 = M$. That is, M is an equivalent similarity matrix, denoted by $\overline{M}$.

With different values of λ, λ-cutting matrix $\overline{M}_\lambda = (\overline{s}_{ij}^{\lambda})_{m\times m}$ of $\overline{M}$ is constructed by Equation (7) and obtain different categories, as present below:

(i) If $0 \le \lambda \le 0.7733, \overline{M}_\lambda = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$,

then the diagnoses are the same category: $\{D, DF, TB\}$.

(ii) If $0.7733 < \lambda \le 0.8666, \overline{M}_\lambda = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, then the diagnoses are the same category: $\{D, DF, TB\}$.

(iii) If $0.8666 < \lambda \le 0.9067, \overline{M}_\lambda = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$, then the diagnoses can be divided into two categories: $\{D\}, \{DF, TB\}$.

(iv) If $0.9067 < \lambda \le 1, \overline{M}_\lambda = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. then the diagnoses can be divided into three categories: $\{D\}, \{DF\}, \{TB\}$.

4.2.2 Clustering analysis using $S_2(A, B)$

With similar steps, by Using $S_2(A, B)$, the similarity matrix is constructed as follows:

$$M = \begin{bmatrix} 1 & 0.6304 & 0.7646 \\ 0.6304 & 1 & 0.8293 \\ 0.7646 & 0.8293 & 1 \end{bmatrix},$$

The identical similarity matrices by limited time compositions of M:

$$M^2 = \begin{bmatrix} 1 & 0.6304 & 0.7646 \\ 0.6304 & 1 & 0.8293 \\ 0.7646 & 0.8293 & 1 \end{bmatrix},$$

It is clearly that $M^2 = M$. That is, M is an equivalent similarity matrix, denoted by $\overline{M}$.

Then, λ-cutting matrix $\overline{M}_\lambda = (\overline{s}_{ij}^{\lambda})_{m\times m}$ of $\overline{M}$ by Equation (7) for different categories as shown below:

(i) If $0 \le \lambda \le 0.6304, \overline{M}_\lambda = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, then the diagnoses are the same category: $\{D, DF, TB\}$.

(ii) If $0.6304 < \lambda \leq 0.7646, \overline{M}_\lambda = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, then the diagnoses are the same category: $\{D, DF, TB\}$.

(iii) If $0.7646 < \lambda \leq 0.8293, \overline{M}_\lambda = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$, then the diagnoses can be divided into two categories: $\{D\}, \{DF, TB\}$.

(iv) If $0.8293 < \lambda \leq 1, \overline{M}_\lambda = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. then the diagnoses can be divided into three categories: $\{D\}, \{DF\}, \{TB\}$.

4.2.3 Clustering analysis using $S_3(A, B)$

By using $S_3(A, B)$, the following similarity matrix is construct:

$$M = \begin{bmatrix} 1 & 0.9165 & 0.9527 \\ 0.9165 & 1 & 0.9638 \\ 0.9527 & 0.9638 & 1 \end{bmatrix}.$$

The equivalent similarity matrices by limited time compositions of M:

$$M^2 = \begin{bmatrix} 1 & 0.9165 & 0.9527 \\ 0.9165 & 1 & 0.9638 \\ 0.9527 & 0.9638 & 1 \end{bmatrix},$$

It is clearly that $M^2 = M$. That is, M is an equivalent similarity matrix, denoted by $\overline{M}$.

λ-cutting matrix $\overline{M}_\lambda = (\overline{s}_{ij}^{\lambda})_{m\times m}$ of $\overline{M}$ is constructed for different categories as shown below:

(i) If $0 \leq \lambda \leq 0.9165, \overline{M}_\lambda = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, then the diagnoses are in the same category: $\{D, DF, TB\}$.

(ii) If $0.9165 < \lambda \leq 0.9527, \overline{M}_\lambda = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, then the diagnoses are in the same category: $\{D, DF, TB\}$.

(iii) If $0.9527 < \lambda \leq 0.9638, \overline{M}_\lambda = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$, then the diagnoses can be divided into two categories: $\{D\}, \{DF, TB\}$.

(iv) If $0.9638 < \lambda \leq 1, \overline{M}_{\lambda} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. then the diagnoses can be divided into three categories: $\{D\}, \{DF\}, \{TB\}$.

4.2.4 Discussions

By utilizing three distance-based similarity measures, $S_1(A, B)$, $S_2(A, B)$ and $S_3(A, B)$, as can be seen that the results of clustering show the same results. In the context of the given medical dataset, four distinct scenarios can be generated through the application of a clustering algorithm that utilizes the proposed three different similarity measures of SVNSs. This indicates that these similarity measures can segment the data into four different clusters, each representing a different classification or pattern within the dataset. The effectiveness of these similarity measures is demonstrated by their capacity to consistently and accurately divide the dataset into meaningful clusters.

Therefore, clustering algorithms incorporating these three similarity measures prove to be reliable tools in clustering problems, particularly in those that involve complex and nuanced data, such as medical datasets. They bring a level of precision and interpretability to the task of grouping similar data points together, facilitating improved data analysis and decision-making processes.

5 CONCLUSION

This project has suggested a clustering method that applies SVNS similarity measures to medical diagnosis. The efficiency of the proposed similarity measure in delivering appropriate and accurate results in case study have been demonstrated and verified through comparisons with previous study. The proposed similarity metric has demonstrated encouraging results when used in medical diagnosis decision-making, showing its potential to improve diagnostic precision and treatment planning. The ability to recognize significant patterns and relationships in medical data is made possible by the ability to capture the inherent uncertainty and ambiguity of SVNS. In conclusion, clustering algorithms can assist in finding complex patterns in medical data that might not be evident or even possible to discover otherwise, leading to improved diagnosis, prevention, and treatment methods in healthcare.

ACKNOWLEDGEMENT

This research was funded by Universiti Teknologi MARA, grant number 600-RMC 5/3/GPM (054/2022). The authors would like to express gratitude to the management of Universiti Teknologi MARA, Cawangan Kelantan and Research Management Centre, Universiti Teknologi MARA for their support.

REFERENCES

[1] L. Zadeh, “Fuzzy sets,” *Information and Control*, vol. 8, pp. 338–353, 1965.

[2] J. S. Chai, G. Selvachandran, F. Smarandache, V. C. Gerogiannis, L. H. Son, Q.-T. Bui, and B. Vo, “New similarity measures for single-valued neutrosophic sets with applications in pattern

recognition and medical diagnosis problems," *Complex & Intelligent Systems*, vol. 7, no. 2, pp. 703–723, 2021.

[3] K. Atanassov, "Intuitionistic fuzzy sets," *International journal bioautomation*, vol. 20, p. 1, 2016.

[4] F. Smarandache, *Neutrosophy: Neutrosophic Probability, Set, and Logic: Analytic Synthesis & Synthetic Analysis*. Rehoboth, NM: American Research Press, 1998.

[5] V. Ulucay, A. Kılıç, M. Sahin, and H. Deniz, *A new hybrid distance-based similarity measure for refined neutrosophic sets and its application in medical diagnosis*. Infinite Study, 2019.

[6] S. Alias and D. Mohamad, "A review on neutrosophic set and its development, discovering mathematics," *Menemui Matematik*, vol. 39, pp. 61–69, 2017.

[7] S. Broumi and F. Smarandache, "Several similarity measures of neutrosophic sets," *Neutrosophic Sets And Systems*, 2013.

[8] H. Wang, F. Smarandache, Y. Zhang, and R. Sunderraman, *Single valued neutrosophic sets*. Infinite study, 2010.

[9] S. Das, B. K. Roy, M. B. Kar, S. Kar, and D. Pamučar, "Neutrosophic fuzzy set and its application in decision making," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 11, pp. 5017–5029, 2020.

[10] H. Yang, Z. L. Guo, Y. She, and X. Liao, "On single valued neutrosophic relations," *Journal of Intelligent & Fuzzy Systems*, vol. 30, no. 2, pp. 1045–1056, 2016.

[11] J. Ye, "Clustering methods using distance-based similarity measures of single-valued neutrosophic sets," *Journal of Intelligent Systems*, vol. 23, no. 4, pp. 379–389, 2014.

[12] H. Alashwal, M. E. Halaby, J. J. Crouse, A. Abdalla, and A. A. Moustafa, "The application of unsupervised clustering methods to alzheimer's disease," *Frontiers in Computational Neuroscience*, vol. 13, 2019.

[13] E. H. Ruspini, "A new approach to clustering," *Information and control*, vol. 15, no. 1, pp. 22–32, 1969.

[14] H. L. Huang, "New distance measure of single-valued neutrosophic sets and its application," *International Journal of Intelligent Systems*, vol. 31, no. 10, pp. 1021–1032, 2016.

[15] P. Majumdar and S. K. Samanta, "On similarity and entropy of neutrosophic sets," *Journal of Intelligent & Fuzzy Systems*, vol. 26, no. 3, pp. 1245–1252, 2014.

[16] S. Pramanik, P. Biswas, and B. C. Giri, "Hybrid vector similarity measures and their applications to multi-attribute decision making under neutrosophic environment," *Neural computing and Applications*, vol. 28, pp. 1163–1176, 2017.

[17] G. Shahzadi, M. Akram, and A. B. Saeid, "An application of single-valued neutrosophic sets in

medical diagnosis," *Neutrosophic sets and systems*, vol. 18, pp. 80–88, 2017.

[18] N. Mustapha, S. Alias, R. M. Yasin, N. N. Mohd Yusof, N. N. Fakhrarazi, and N. N. Aisyah Nik Hassan, "New entropy measure concept for single value neutrosophic sets with application in medical diagnosis." *International Journal of Neutrosophic Science (IJNS)*, vol. 19, no. 1, 2022.

[19] S. Ye, J. Fu, and J. Ye, "Medical diagnosis using distance-based similarity measures of single valued neutrosophic multisets," *Neutrosophic Sets and Systems*, vol. 7, no. 1, pp. 47–52, 2015.

[20] K. Mondal and S. Pramanik, "Intuitionistic fuzzy similarity measure based on tangent function and its application to multi-attribute decision making," *Global journal of advanced research*, vol. 2, no. 2, pp. 464–471, 2015.

[21] M. Luo and R. Zhao, "A distance measure between intuitionistic fuzzy sets and its application in medical diagnosis," *Artificial Intelligence in Medicine*, vol. 89, pp. 34–39, 2018.

[22] F. Xiao, "A hybrid fuzzy soft sets decision making method in medical diagnosis," *IEEE Access*, vol. 6, pp. 25 300–25 312, 2018.

[23] N. Mustapha, S. Alias, R. M. Yasin, I. Abdullah, and S. Broumi, "Cardiovascular diseases risk analysis using distance-based similarity measure of neutrosophic sets," *Neutrosophic Sets and Systems*, vol. 47, pp. 26–37, 2021.

[24] R. Lowen, A. Jones, A. Kaufmann, and H. Zimmermann, *Fuzzy Sets Theory and Applications*. Springer Netherlands, 1986.