Computer Systems Science & Engineering
DOI:10.32604/csse.2021.015628
Article

Tech Science Press

Generalized Normalized Euclidean Distance Based Fuzzy Soft Set Similarity for Data Classification

Rahmat Hidayat[1,2,*], Iwan Tri Riyadi Yanto[1,3], Azizul Azhar Ramli[1], Mohd Farhan Md. Fudzee[1] and Ansari Saleh Ahmar[4]

[1]Faculty of Computer and Information Technology, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Malaysia
[2]Department of Information Technology, Politeknik Negeri Padang, Padang, Indonesia
[3]Department of Information System, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[4]Department of Statistics, Universitas Negeri Makassar, Makassar, Indonesia
[*]Corresponding Author: Rahmat Hidayat. Email: rahmat@pnp.ac.id
Received: 30 November 2020; Accepted: 17 February 2021

Abstract: Classification is one of the data mining processes used to predict predetermined target classes with data learning accurately. This study discusses data classification using a fuzzy soft set method to predict target classes accurately. This study aims to form a data classification algorithm using the fuzzy soft set method. In this study, the fuzzy soft set was calculated based on the normalized Hamming distance. Each parameter in this method is mapped to a power set from a subset of the fuzzy set using a fuzzy approximation function. In the classification step, a generalized normalized Euclidean distance is used to determine the similarity between two sets of fuzzy soft sets. The experiments used the University of California (UCI) Machine Learning dataset to assess the accuracy of the proposed data classification method. The dataset samples were divided into training (75% of samples) and test (25% of samples) sets. Experiments were performed in MATLAB R2010a software. The experiments showed that: (1) The fastest sequence is matching function, distance measure, similarity, normalized Euclidean distance, (2) the proposed approach can improve accuracy and recall by up to 10.3436% and 6.9723%, respectively, compared with baseline techniques. Hence, the fuzzy soft set method is appropriate for classifying data.

Keywords: Soft set; fuzzy soft set; classification; normalized euclidean distance; similarity

1 Introduction

Nowadays, Big Data is used in Tuberculosis (TBC) patient data in healthcare, stock data in economics and business fields, and BMKG data (containing weather, temperature, and rainfall data), etc. Data mining is the process of extracting knowledge from large amounts of data [1], and is done by extracting information and analyzing data patterns or relationships [2,3].

This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Classification is one of the data mining processes used to predict predetermined target classes with data learning accurately. The classification has been used in health [4–6], economics, and agriculture fields [7,8]. Classifying data is challenging and requires further research [9].

In 1965, Zadeh [10] introduced a fuzzy set in which each element object had a grade of memberships ranging between zero and one. In comparison, Molodtsov [11] introduced soft set theory to collect parameters from the universal set subsets (set U). Soft set theory is widely used to overcome the presence of elements of uncertainty or doubt, such as those found in decision-making. Roy developed fuzzy soft set theory by combining soft set theory and fuzzy set theory. This theory was then used in decision-making problems [12,13]. Majumdar and Samanta [14] presented a fuzzy soft set for similarity measurement between two generalized fuzzy soft sets for decision-making.

The fuzzy soft set, an extension of the classical soft set, was introduced by Maji [15]. There have been many works about fuzzy soft set theory in decision-making. Ahmad et al. [16] defined arbitrary fuzzy soft union and fuzzy soft intersection and proved Demorgan laws using fuzzy soft set theory. Meanwhile, Aktas and Cagman [17] studied fuzzy parameterized soft set theory, related properties, and decision-making applications. Rehman et al. [18] studied some fuzzy soft sets' operations and gave fuzzy soft sets the fundamental properties. Finally, Celik et al. [19] researched applications of fuzzy soft sets in ring theory.

The critical issue in fuzzy soft sets is the similarity measure. In recent years, similarity measurement between two fuzzy soft sets has been studied from different aspects and applied to various fields, such as decision-making, pattern recognition, region extraction, coding theory, and image processing. For example, similarity measurement [20] has been researched in fuzzy soft sets based on distance, set-theoretic approaches, and matching functions. Sut [21] and Rajarajeswari [22] used the notion of the similarity measure in Majumdar and Samanta [20] to make decisions. Several similarity measurement [23] based on four types of quasi-metrics were introduced to fuzzy soft sets. Sulaiman [24] researched a set-theoretic similarity measure for fuzzy soft sets, and applied it to group decision-making. However, some studies haphazardly investigated the similarity measurement of fuzzy soft sets based on distance, resulting in high computational costs [20,23]. Feng and Zheng [25] showed that the similarity measure based on the Hamming distance and normalized Euclidean distance in the fuzzy soft set is reasonable. Thus, the similarity of generalized normalized Euclidean distance is applied in the present paper to a fuzzy soft set for classification. The similarity is used to classify the label of data. The experimental results show that the proposed approach can improve classification accuracy.

2 The Proposed Method/Algorithm

This section presents the basic definitions of fuzzy set theory, soft set theory, and some useful definitions from Roy and Maji [12].

2.1 Fuzzy Set

Definition 2.1 [10] Let U be a universe. A fuzzy set A over U is a set defined by a function

$$\mu_A : U \rightarrow [0, 1] \tag{1}$$

where μ_A is the membership function of A, and the value $\mu_A(x)$ is the membership value of $x \in U$. The value represents the degree of x belonging to the fuzzy set U. Thus, a fuzzy set A over U can be represented as in (2).

$$A = \{\mu_A(x) \mid x \in U, \mu_A(x) \in [0, 1]\} \tag{2}$$

The notion that the set of all the fuzzy sets over U was denoted by $F(U)$.

Definition 2.2 [10] Let A be a fuzzy set, where $A \in F(U)$. Then, the complement of A is as in (3)

$$A^c = \{\mu_A{}^c(x) \mid x \in U, \mu_A{}^c(x) = 1 - \mu_A(x)\} \tag{3}$$

Definition 2.3 [10] Let A, B be the fuzzy set, where A, $B \in F(U)$. The membership degree of union of A and B is denoted by $\mu A \cup B(x)$:

$$\mu A \cup \mathrm{B}(\mathrm{x}) = \max\{\mu_{\mathrm{A}}(\mathrm{x}),\ \mu_{\mathrm{B}}(\mathrm{x})\}; \tag{4}$$

for all $x \in U$ and $\mu A \cup B(x) \in [0,1]$.

Definition 2.4 [10] Let A,B be the fuzzy set, where A,$B \in F(U)$. The membership degree of intersection of A and B is denoted by $\mu_{A\cap B}(x)$:

$$\mu_{A\cap B}(x) = min\{\mu_A(x),\ \mu_B(x)\}; \tag{5}$$

for all $x \in U$ and $\mu A \cup B(x) \in [0,1]$.

2.2 Fuzzification

Fuzzification is a process that changes the crisp value to a fuzzy set, or a fuzzy quantity into a crisp quantity [26]. This process uses the membership function and fuzzy rules. The fuzzy rules can be formed as fuzzy implications, such as (x_1 *is* A_1) ° (x_2 *is* A_2) ° … ° (x_n *is* A_n); then Y is B, with ° being the operator "AND" or "OR". B can be determined by combining all antecedent values [14].

2.3 Fuzzy Soft Set (FSS)

Definition 2.5 [12] Let U be an initial universe set and E be a set of parameters. Let $P(U)$ denote the power set of all fuzzy subsets of U, and $A \subseteq E$. Γ_A is called a fuzzy soft set over U, where the function of γ_A is a mapping given by $\gamma_A : A \to P(U)$ such that $\gamma_A(e) = \varnothing$ if $e \notin A$.

Here, the function γ_A is an approximate function of the fuzzy soft set Γ_A, and the value $\gamma_A(e)$ is called an e-element of a fuzzy soft set for all $e \notin A$. Fuzzy soft set Γ_A over U can be represented by the set of ordered pairs:

$$\Gamma_A = \{(e, \gamma_A(e))|e \in A, \gamma_A(e) \in P(U)\}. \tag{6}$$

Note that the set of all the fuzzy soft sets over U was denoted by $FS(U)$.

Example 1 [14] Let a fuzzy soft set Γ_A describe the attractiveness of the shirt concerning the given parameters, which the authors are going to wear. $U = \{u_1, u_2, u_3, u_4, u_5\}$ is the set of all shirts under consideration. $P(U)$ be the collection of all fuzzy subsets of U. Let E = {e_1 = "colorful", e_2 = "bright", e_3 = "cheap", e_4 = "warm"}. If $A = \{e_1, e_2, e_3\}$ can be the approximate value of the function fuzzy,

γ_A (e_1) = {0.5|u_1, 0.9|u_2},

γ_A (e_2) = {1|u_1, 0.8|u_2, 0.7|u_3},

γ_A (e_3) = {1|u_2, 1|u_5}.

The family {γ_A (e_i); i = 1,2,3} of *P(U)* is then a fuzzy soft set Γ_A. The tabular representation for fuzzy soft set Γ_A is shown in Tab. 1.

Definition 2.6 [14] Let Γ_A, $\Gamma_B \in FS(U)$. Γ_A is a fuzzy soft subset of Γ_B, denoted by $\Gamma_A \subseteq \Gamma_B$, if γ_A(e) $\subseteq$ γ_B(e) for all $e \in A$, $A \subseteq B$.

Definition 2.7 [14] Let Γ_A, $\in FS(U)$. The complement of fuzzy soft set Γ_A is denoted by $\Gamma_A{}^c$ such that $\gamma_{A^c}(e) = \gamma_A^c(e)$ for all $e \in A$.

Table 1: The representation of the fuzzy soft set Γ_A

U/A	e_1	e_2	e_3
x_1	0.5	1	0
x_2	0.9	0.8	1
x_3	0	0.7	0
x_4	0	0	0
x_5	0	0	0.3

Definition 2.8 [14] Let $\Gamma_A, \Gamma_B \in FS(U)$. The union of Γ_A and Γ_B is denoted by $\Gamma_{A \cup B}(e) = \gamma_A(e) \cup \gamma_B(e)$ for all $e \in A \cup B$ $e \in A \cup B$.

Definition 2.9 [14] Let $\Gamma_A, \Gamma_B \in FS(U)$. The intersection of Γ_A and Γ_B is denoted by $\Gamma_{A \cap B}(e) = \gamma_A(e) \cup \gamma_B(e)$ for all $e \in A \cup B$ $e \in A \cap B$.

Definition 2.10 [14] Let $\Gamma_A, \in FS(U)$. The cardinal set of Γ_A, denoted by $c\Gamma_A$, can be defined by $c\Gamma_A = \{\mu\ c\Gamma_A(e) | e : e \in A\}$, where membership function $\mu\ c\Gamma_A$ of $c\Gamma_A$ is defined by

$$c\Gamma_A : E \to [0, 1] \tag{7}$$

$$\mu_{c\Gamma_A}(e) = \frac{|\mu_A(e)|}{|U|}. \tag{8}$$

$|U|$ is the cardinality of universe U, and

$$|\mu_A(e)| = \sum\nolimits_{u \in U} \mu_{\gamma_A(e)}(u). \tag{9}$$

The set of all cardinal sets of fuzzy soft set over U can be denoted by $cFS(U)$.

2.4 Classification

Classification involves learning a target function that maps each collection of data attributes to several groups of predefined classes. The purpose of the classification is to see the class's target predictions as accurate as possible for each case in the data. The classification algorithm consists of two stages. In the training stage, the classifier is trained on predefined classes or data categories. An X tuple, represented by the n-dimensional vector attribute, $X = \{x_1, x_2, \ldots, x_N\}$, describes by the measurements made on the tuples with n attributes $A_1, A_2, \ldots, A_M$. Each tuple belongs to a class, as identified by its attributes. Class attribute labels have discreet, non-consecutive values, and each value acts as a category or class. Next, the second step is Classification. In this step, the built-in classifier was used to classify the data by looking at the classification algorithm's accuracy in the estimated data testing. The step is to see the accuracy in the first classification; the predicted classifier's accuracy is estimated. If using a training set to measure the classifier's accuracy, then the estimate would be optimal because the data used to form the classifier comprise the training set. Therefore, a test set (a set of tuples and their class labels selected randomly from the dataset) were used. Test sets are independent of the training sets because test sets were not used to build a classifier.

2.5 Similarity Measurement

A measurement of similarity or dissimilarity defines the relationships between samples or objects. Similarity measurements were used to determine which patterns, signals, images, or sets are alike. For the

similarity measure, the resemblance is more critical when its value increases, but, conversely, for a dissimilarity measurement, the resemblance is more robust when its value decreases [27]. An example of the dissimilarity measure is a distance measure. Measuring similarity or distance between two entities is crucial in various data mining and information discovery tasks, such as classification and clustering. Similarity indicators calculate the degree that various patterns, signals, images, or sets are alike. A few researchers have measured the similarity between fuzzy sets, fuzzy numbers, and vague sets. Recently [14,20,28] studied the similarity measure of the soft set and fuzzy soft set. They explained the similarity between the two generalized fuzzy soft sets as follows.

Let $U = \{x_1, x_2, \ldots, x_n\}$ be the universal set of elements and $E = \{e_1, e_2, \ldots, e_m\}$ be the universal set of parameters. Let $F\rho$ and $G\delta$ be two generalized fuzzy soft sets over the parameterized universe (U, E). Hence, $F\rho = \{F(e_i), \rho(e_i), i = 1, 2, \ldots, m\}$ and $G\delta = \{G(e_i), \delta(e_i), i = 1, 2, \ldots, m\}$. Thus, $F = \{F(e_i),\ i = 1, 2,\ \ldots,\ m\}$ and $G = \{G(e_i),\ i = 1, 2,\ \ldots,\ m\}$ are two families of fuzzy soft sets.

The similarity between F and G is found and denoted by M(F,G). Next, the similarity between the two fuzzy sets ρ and δ is found and denoted by m (ρ,δ). Then, the similarity between the two generalized fuzzy soft sets $F\rho$ and $G\delta$ is denoted as S($F\rho$,$G\delta$) = M(F,G) $\times$ m(ρ,δ).

Therefore, M (F, G) = max M_i (F,G), where:

$$M_i(F_-, G_-) = 1 - \frac{\sum_{j=1}^{n} |F_{-ij} - G_{-ij}|}{\sum_{j=1}^{n} (F_{-ij} + G_{-ij})}. \tag{10}$$

Furthermore,

$$m(\rho, \delta) = 1 - \frac{\sum_{j=1}^{n} |\rho_i - \delta_i|}{\sum_{j=1}^{n} (\rho_i + \delta_i)}. \tag{11}$$

If we use the universal fuzzy soft set, then $\rho = \delta = 1$ and m(ρ,δ) = 1. Now, the formula for similarity is

$$S(F_\rho, G_\delta) = M_i(F_-, G_-) = 1 - \frac{\sum_{j=1}^{n} |F_{-ij} - G_{-ij}|}{\sum_{j=1}^{n} (F_{-ij} + G_{-ij})}. \tag{12}$$

Example 2. In this example, $U = \{x_1, x_2, x_3, x_4\}$ and $E = \{e_1, e_2, e_3\}$. Let there be two generalized fuzzy soft sets over the parameterized universe (U, E).

Here,

$$m(\rho, \delta) = 1 - \frac{\sum_{i=1}^{3} |\rho_i - \delta_i|}{\sum_{i=1}^{3} (\rho_i + \delta_i)} = 1 - \frac{0.1 + 0.1 + .05}{1.1 + 1.5 + 1.3} = 0.82$$

and $M_1(F,G) \cong 0.73$; $M_2(F,G) \cong 0.43$; $M_3(F,G) \cong 0.50$. Thus, max [$M_i(F,G)$] $\cong 0.73$.

Hence, the similarity between the two GFSS $F\rho$ and $G\delta$ were S($F\rho$,$G\delta$) = M(F,G) $\times$ m(ρ,δ) = 0.73 $\times$ 0.82 = 0.60 for a universal fuzzy soft set, where $\rho = \delta = 1$ and m(ρ,δ) = 1. Then, the similarity S ($F\rho$,$G\delta$) = 0.73.

2.6 Distance Measurement

In this study, the fuzzy soft set was calculated based on the normalized Hamming distance [25]. We assume fuzzy soft sets (F,A) and (G,B) have the same set of parameters, namely, A = B. The normalized Hamming distance and normalized distance in Fuzzy Soft Set (FSS) are obtained using Eqs. (13) and (14).

$$d_1((F,A),(G,B)) = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left|F(e_i)(x_j) - G(e_i)(x_j)\right| \tag{13}$$

$$d_2((F,A),(G,B)) = \frac{1}{mn}\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\left|F(e_i)(x_j) - G(e_i)(x_j)\right|^2\right)^{\frac{1}{2}} \tag{14}$$

Example 3. As in Roy and Maji [12], let U = {u$_1$, u$_2$, u$_3$} be a set with parameters $= \{a_1,\ a_2, a_3\}$. Two FSS (G,A) and (H,A) are represented by Tabs. 2 and 3, respectively.

Table 2: Fuzzy set (G,A)

(G,A)	a_1	a_2	a_3
u_1	0.7	0.8	0.6
u_2	0.6	0.7	0.5
u_3	0.5	0.8	0.8

Table 3: Fuzzy set (H,A)

(H,A)	a_1	a_2	a_3
u_1	0.5	0.6	0.9
u_2	0.7	0.8	0.6
u_3	0.4	0.8	1

Using Eqs. (13) and (14), respectively, the normalized Hamming distance and normalized distance in FSS between (G,A) and (H,A) can be calculated as follows:

$$d_1((G,A),(H,A)) = \frac{1}{3\times 3}\sum_{i=1}^{3}\sum_{j=1}^{3}(0.2+0.1+0.1+0.2+0.1+0+0.3+0.1+0.2) \approx 0.144$$

and

$$d_2((F,E),(G,E)) = \frac{1}{3\times 3}\sum_{i=1}^{3}\sum_{j=1}^{3}\left(0.2^2+0.1^2+0.1^2+0.2^2+0.1^2+0^2+0.3^2+0.1^2+0.2^2\right)^{\frac{1}{2}} \approx 0.056$$

Feng and Zheng [13] extended Eq. (14) into a generalized normalized distance in FSS:

$$d_4((F,A),(G,B)) = \frac{1}{m}\sum_{i=1}^{m}[\frac{1}{n}\sum_{j=1}^{n}|F(e_i)(x_j) - G(e_i)(x_j)|^p)^{\frac{1}{p}}],\ p\in N_+. \tag{15}$$

If $p = 1$, then Eq. (13) is reduced to Eq. (14).

From Eq. (14), it can be known that

$$d' = \frac{1}{n}\sum_{j=1}^{n}\left|F(e_i)(x_j) - G(e_i)(x_j)\right| \tag{16}$$

d' indicates the distance between the i^{th} parameter of (F,A) and (G,B), and $d_1((F,A),(G,B))$ indicates the distance among all parameters of (F,A) and (G,B).

3 Discussion

In this section, the proposed approach and experimental results of the Fuzzy Soft Set Classifier (FSSC) using the normalized Euclidean distance are discussed.

3.1 Proposed Approach

This study proposed a new classification algorithm based on the fuzzy soft set; we call it the Fuzzy Soft Set Classifier (FSSC). This algorithm used the normalized Euclidean distance of similarity between two fuzzy soft sets to classify unlabeled data. Before training and classification steps, we first conducted fuzzification and created a fuzzy soft set.

3.1.1 Training Step

The goal of training the algorithm is to determine the center of each existing class.

Let $U = \{u_1, u_2, \ldots, u_N\}$, E be the set of parameters, $A \subseteq E$, *and* $A = \{e_i, i = 1, 2, \ldots M\}$. There are k classes with n_r samples in each class, where $r = 1, 2, \ldots, k$ and $\sum_{r=1}^{k} n_r = N$. Let us say that $C_r \subseteq U$ is r-class data, and $\Gamma_{C_r,}$ is the set of fuzzy soft sets of the r-class data. Thus, the center set of class C_r is denoted as Γ_{PC_r} and be defined as in Eq. (17).

$$\Gamma_{PC_r} = c\,\Gamma_{C_r} = \mu c\,\Gamma_{C_r}(e_i) = \frac{\gamma_{C_r}(e_i)}{|C_r|} = \frac{\sum_{j=1}^{n_r} \mu_{\gamma_{C_r}(e_i)}(u_j)}{n_r}$$

Thus,

$$\Gamma_{PC_r} = \frac{1}{n_r}\sum\nolimits_{j=1}^{n_r} \mu_{\gamma_{C_r}(e_i)}(u_j),\ \forall e_i, i = 1, 2, \ldots, m, \forall C_r, r = 1, 2, \ldots, k \tag{17}$$

3.1.2 Classification Step

The new data of the training step results were used to determine the classes in the new data; that is, by measuring the similarity of two sets of fuzzy soft sets acquired in the class center vector and new data.

Given $\Gamma_{C_r}, r = 1, 2, \ldots, k$ fuzzy soft set of new data Γ_G. The formula for measuring similarity:

similarity measure $= 1 -$ *disctance measure*.

We use the generalized normalized Euclidean distance for normalized Euclidean distance of the fuzzy set. With relation to Eq. (15), rather than the normalized Euclidean distance fuzzy set,

$$q(A, B) = \frac{1}{m}\sum\nolimits_{i=1}^{m} [\frac{1}{n}\sum\nolimits_{j=1}^{n} |F(e_i)(x_j) - G(e_i)(x_j)|^p)^{\frac{1}{p}}],\ p \in N_+. \tag{18}$$

The generalized normalized Euclidean distance fuzzy soft set is as follows:

$$Q * (\Gamma_{P_{C_r}}, \Gamma_G) = \left(\frac{1}{m.n}\left(\sum\nolimits_{i=1}^{m}\sum\nolimits_{j=1}^{n_r}\left(\gamma_{P_{C_r}}(e_i)(x_j) - \gamma_G(e_i)(x_j)\right)^p\right)\right)^{\frac{1}{p}}, \tag{19}$$

$$\Leftrightarrow Q * (\Gamma_{P_{C_r}}, \Gamma_G) = \left(\frac{1}{m.1}\left(\sum\nolimits_{i=1}^{m}\left(\gamma_{P_{C_r}}(e_i)(x_1) - \gamma_G(e_i)(x_1)\right)^p\right)\right)^{\frac{1}{p}}, \tag{20}$$

$$\Leftrightarrow Q * (\Gamma_{P_{C_r}}, \Gamma_G) = \left(\frac{1}{m}\left(\sum\nolimits_{i=1}^{m}\left(\gamma_{P_{C_r}}(e_i)(x) - \gamma_G(e_i)(x)\right)^p\right)\right)^{\frac{1}{p}}. \tag{21}$$

Thus, the formula for the similarity measure becomes:

$$S * \left(\Gamma_{P_{C_r}}, \Gamma_G\right) = 1 - Q * \left(\Gamma_{P_{C_r}}, \Gamma_G\right), \tag{22}$$

$$\Leftrightarrow S * \left(\Gamma_{P_{C_r}}, \Gamma_G\right) = 1 - \left(\frac{1}{m}\left(\sum_{i=1}^{m}\left(\gamma_{P_{C_r}}(e_i)(x) - \gamma_G(e_i)(x)\right)^p\right)\right)^{\frac{1}{p}}. \tag{23}$$

After the value the similarity for each class was obtained, the algorithm looked for which class label is appropriate for new data Γ_G by determining the maximum similarity.

$$prediction = arg\left[\max_{r=1}^{k} S * \left(\Gamma_{P_{C_r}}, \Gamma_G\right)\right]. \tag{24}$$

3.2 Experimental Results

We conducted experiments using the University of California (UCI) dataset to assess the accuracy of the proposed data classification method. The dataset samples were divided into training (75% of samples) and test (25% of samples) sets. Experiments were performed in MATLAB R2010a software. Figs. 1–4 show the classification results obtained by our fuzzy soft set method and other baseline techniques.

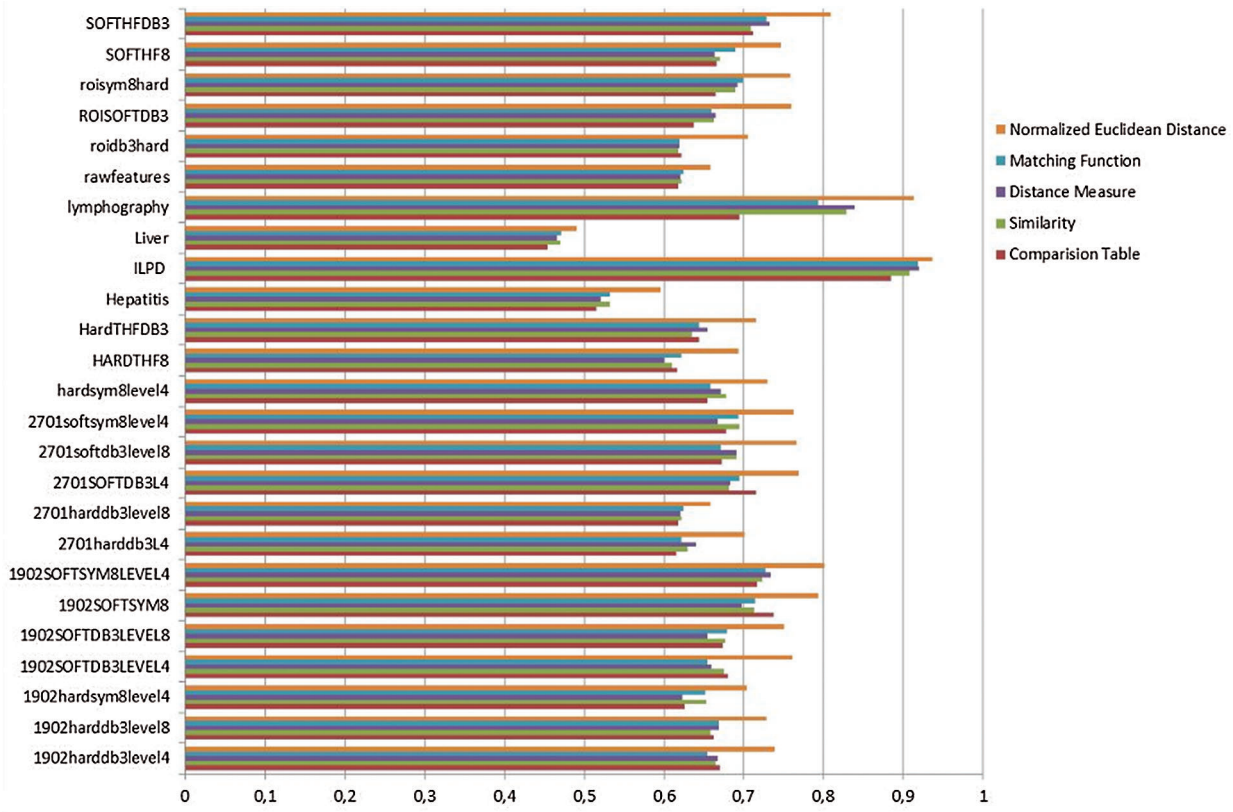

Figure 1: Comparison of accuracy

As seen in Fig. 1, calculations using the normalized Euclidean distance method yield the highest accuracy results. Fig. 2 shows that the normalized Euclidean distance method obtains the second-highest precision; the highest precision is obtained by the comparison table method in MatLab.

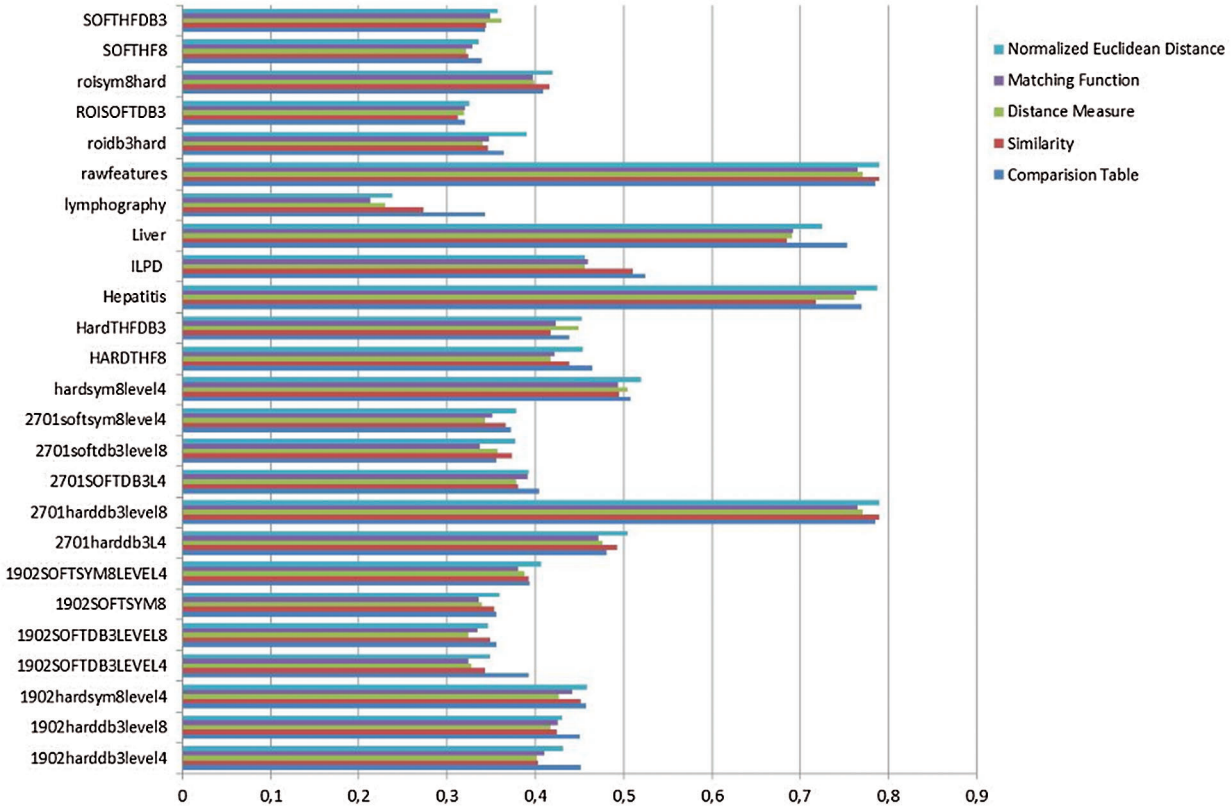

Figure 2: Comparison of precision

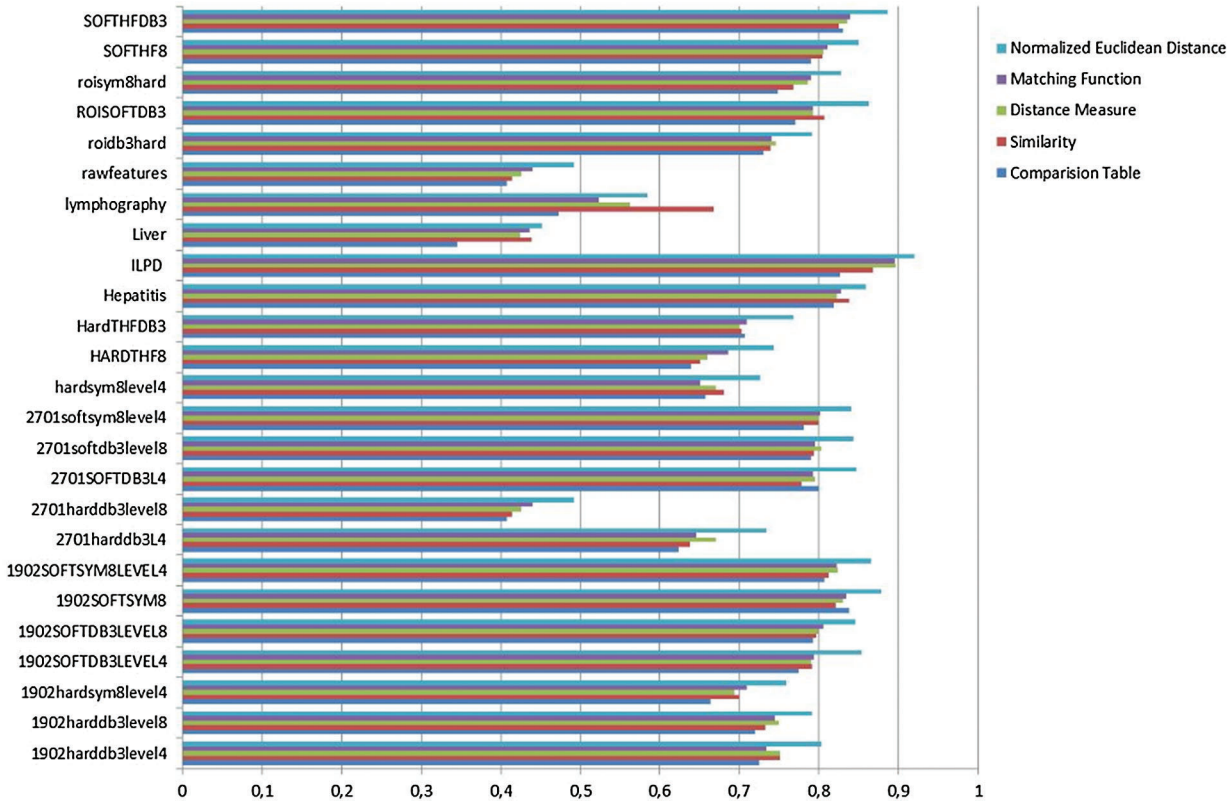

Figure 3: Comparison of recall

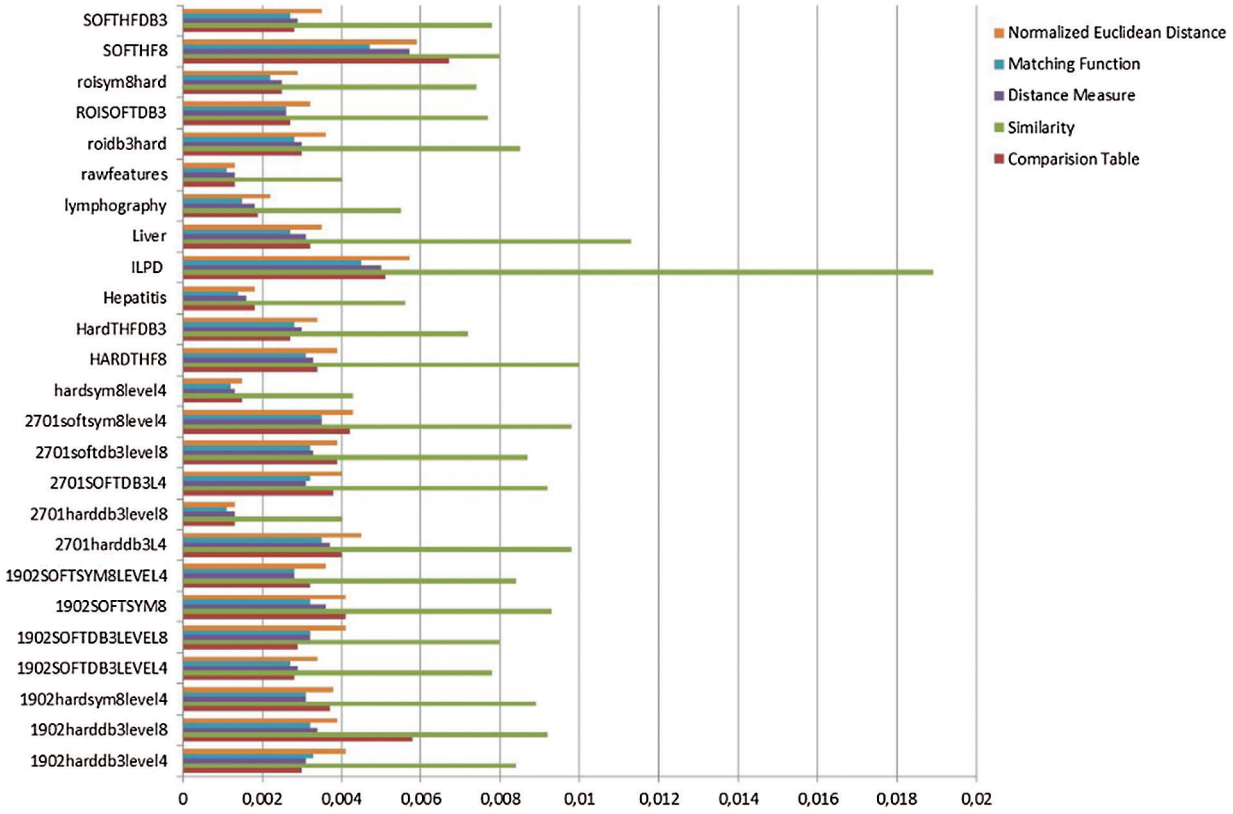

Figure 4: Comparison of computational time

Fig. 3 shows that the normalized Euclidean distance method produces the highest recall results, whereas Fig. 4 illustrates that the method has the highest computation time.

The fastest sequence is matching function, distance measure, similarity, normalized Euclidean distance. Comparisons are shown in Tab. 4.

Table 4: Improvement of accuracy and recall

	Comparison Table	Similarity	Distance Measure	Matching Function	Normalized Euclidean Distance	Improvement
Accuracy	0.6580	0.6688	0.6671	0.6689	0.7380	10.3436 %
Recall	0.6986	0.7212	0.7221	0.7222	0.7725	6.9723 %

4 Conclusions

In this study, a new classification algorithm based on fuzzy soft set theory was proposed. Experimental results show that the normalized Euclidean distance method improves accuracy by 10.3436% and increases by 6.9723%, compared to baseline techniques. We also find that all similarity measurements proposed in this paper are reasonable.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no interest in reporting regarding the present study.

References

[1] J. Han, M. Kamber and J. Pei, "13 - Data mining trends and research rontiers BT - data mining," In: J. Han (ed.), *The Morgan Kaufmann Series in Data Management Systems*, 3rd edition, Boston: Morgan Kaufmann, pp. 585–631, 2012.

[2] Y. Cheng, K. Chen, H. Sun, Y. Zhang and F. Tao, "Data and knowledge mining with big data towards smart production," *Journal of Industrial Information Integration*, vol. 9, no. 9, pp. 1–13, 2018.

[3] M. Azarafza, M. Azarafza and H. Akgün, "Clustering method for spread pattern analysis of corona-virus (COVID-19) infection in Iran," *Journal of Applied Science, Engineering, Technology, and Education*, vol. 3, no. 1, pp. 1–6, 2021.

[4] D. E. Lumsden, H. Gimeno and J.-P. Lin, "Classification of dystonia in childhood," *Parkinsonism & Related Disorders*, vol. 33, pp. 138–141, 2016.

[5] M. Zheng, "Classification and pathology of lung cancer," *Surgical Oncology Clinics*, vol. 25, no. 3, pp. 447–468, 2016.

[6] A. Ojugo and O. D. Otakore, "Forging an optimized bayesian network wodel with selected parameters for detection of the coronavirus in Delta State of Nigeria," *Journal of Applied Science, Engineering, Technology, and Education*, vol. 3, no. 1, pp. 37–45, 2021.

[7] X. Li and Y. Tang, "Two-dimensional nearest neighbor classification for agricultural remote sensing," *Neurocomputing*, vol. 142, no. 10–12, pp. 182–189, 2014.

[8] Y. Tang and X. Li, "Set-based similarity learning in subspace for agricultural remote sensing classification," *Neurocomputing*, vol. 173, no. 10–12, pp. 332–338, 2016.

[9] B. Handaga, T. Herawan and M. M. Deris, "FSSC: An algorithm for classifying numerical data using fuzzy soft set theory," *International Journal of Fuzzy System Applications (IJFSA)*, vol. 2, no. 4, pp. 29–46, 2012.

[10] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[11] D. Molodtsov, "Soft set theory—first results," *Computers & Mathematics with Applications*, vol. 37, no. 4–5, pp. 19–31, 1999.

[12] A. R. Roy and P. K. Maji, "A fuzzy soft set theoretic approach to decision making problems," *Journal of Computational and Applied Mathematics*, vol. 203, no. 2, pp. 412–418, 2007.

[13] P. K. Maji, A. R. Roy and R. Biswas, "An application of soft sets in a decision making problem," *Computers & Mathematics with Applications*, vol. 44, no. 8–9, pp. 1077–1083, 2002.

[14] P. Majumdar and S. K. Samanta, "Generalised fuzzy soft sets," *Computers & Mathematics with Applications*, vol. 59, no. 4, pp. 1425–1432, 2010.

[15] P. K. Maji, R. Biswas and A. R. Roy, "Fuzzy soft sets," *Journal of Fuzzy Mathematics*, vol. 9, no. 3, pp. 589–602, 2001.

[16] B. Ahmad and A. Kharal, "On fuzzy soft sets," *Advances in Fuzzy Systems*, vol. 2009, pp. 586507, 2009.

[17] H. Aktaş and N. Çağman, "Soft sets and soft groups," *Information Sciences*, vol. 177, no. 13, pp. 2726–2735, 2007.

[18] A. Rehman, S. Abdullah, M. Aslam and M. S. Kamran, "A study on fuzzy soft set and its operations," *Annals of Fuzzy Mathematics and Informatics*, vol. 6, no. 2, pp. 339–362, 2013.

[19] Y. Celik, C. Ekiz and S. Yamak, "Applications of fuzzy soft sets in ring theory," *Annals of Fuzzy Mathematics and Informatics*, vol. 5, no. 3, pp. 451–462, 2013.

[20] P. Majumdar and S. K. Samanta, "On similarity measures of fuzzy soft sets," *International Journal of AdvanceSoft Computing and Applications*, vol. 3, no. 2, pp. 1–8, 2011.

[21] D. K. Sut, "An Application of similarity of fuzzy soft sets in decision making," *Computer Technology and Application*, vol. 3, no. 2, pp. 742–745, 2012.

[22] D. P. Rajarajeswari and P. Dhanalakshmi, "An application of similarity measure of fuzzy soft set based on distance," *IOSR Journal of Mathematics*, vol. 4, no. 4, pp. 27–30, 2012.

[23] H. Li and Y. Shen, "Similarity measures of fuzzy soft sets based on different distances," in *2012 Fifth International Symposium on Computational Intelligence and Design. Proceedings: IEEE Computer Society (IEEE, 6401247)*. Vol. 1. Hangzhou, China, pp. 527–529, 2012.

[24] N. H. Sulaiman and D. Mohamad, "A set theoretic similarity measure for fuzzy soft sets and its application in group decision making," in *20th National Symposium on Mathematical Sciences: Research in Mathematical Sciences: A Catalyst for Creativity and Innovation. Proceedings: AIP Conference*, Putrajaya, Malaysia, vol. 1522, pp. 237–244, 2012.

[25] Q. Feng and W. Zheng, "New similarity measures of fuzzy soft sets based on distance measures," *Annals of Fuzzy Mathematics and Informatics*, vol. 7, no. 4, pp. 669–686, 2014.

[26] L. Baccour, A. M. Alimi and R. I. John, "Some notes on fuzzy similarity measures and application to classification of shapes, recognition of arabic sentences and mosaic," *IAENG International Journal of Computer Science*, vol. 41, no. 2, pp. 81–90, 2014.

[27] S. Chowdhury and R. Kar, "Evaluation of approximate fuzzy membership function using linguistic input-an approached based on cubic spline," *JINAV: Journal of Information and Visualization*, vol. 1, no. 2, pp. 53–59, 2020.

[28] P. Majumdar and S. K. Samanta, "Similarity measure of soft sets," *New Mathematics and Natural Computation*, vol. 04, no. 01, pp. 1–12, 2008.