Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 10(9): 21994–22022.
DOI: 10.3934/math.2025979
Received: 12 August 2025
Revised: 08 September 2025
Accepted: 12 September 2025
Published: 22 September 2025

Research article

Soft set theory applications: Time series as soft sets and ratio based similarity measure on soft sets

Nazan Polat*

Department of Mathematics, Faculty of Science, Muğla Sıtkı Koçman University, Muğla, Türkiye

* **Correspondence:** Email:ncakmak@mu.edu.tr.

Abstract: Soft set theory, originally introduced by Molodtsov in 1999, offers a versatile and flexible mathematical framework for handling uncertainties, vagueness, and incomplete information in complex systems. Over the past two decades, it has found extensive applications in diverse domains such as decision making, data analysis, engineering, and medical diagnosis. However, despite this wide applicability, its potential in the field of time series analysis remains relatively unexplored and underutilized. Time series data, which capture the evolution of phenomena over time, often involve uncertainty, noise, and nonstationary pattern features that make it an ideal candidate for soft set based modeling. Building on this, we propose two distinct approaches, amplitude threshold soft set and time derivative threshold soft set for representing time series data within the soft set framework. The effectiveness of the proposed methodology is demonstrated through two comprehensive real world applications. First, 58 year temperature dataset from six Brazilian cities spanning from 1967 to 2019 is analyzed, demonstrating how the soft set framework can capture long-term climatological patterns and facilitate analysis across different geographical regions. Second, we consider heart sound classification using phonocardiogram (PCG) data, showing how soft set based time series representation can effectively distinguish between normal and abnormal heart sounds with promising classification performance with the help of a novel ratio based similarity measure specifically designed for soft sets. The methodology's effectiveness is demonstrated through medical signal processing applications, where it achieved competitive performance on heart sound classification (challenge score: 0.776). By integrating soft set theory with time series analysis, and applying the proposed similarity measure in this context, the study bridges an existing gap between the two fields. This approach offers a fresh perspective for pattern recognition, comparative analysis, and uncertainty modeling in temporal data, opening new avenues for future research.

Keywords: soft sets; similarity measure of soft sets; amplitude threshold soft set; time derivative threshold soft set; ratio based similarity measure on soft sets
Mathematics Subject Classification: 03E75, 03E72, 68Q01

1. Introduction

The mathematical foundations of soft set theory were established by Molodtsov [1], who introduced this framework as a novel approach to handle uncertainty without the restrictions imposed by traditional set theories. The fundamental advantage of soft sets lies in their ability to model vague concepts without requiring precise membership functions or probability distributions. Building upon Molodtsov's pioneering work, Maji et al. [2] provided comprehensive theoretical developments and operational definitions that became the cornerstone of subsequent research. The integration of soft sets with fuzzy logic was later explored by Maji et al. [3], introducing fuzzy soft sets that combine the flexibility of soft sets with the graduated membership characteristics of fuzzy sets.

The development of effective similarity measures has been crucial for the practical application of soft sets. Tversky [4] laid the psychological and mathematical foundations for understanding similarity through feature based approaches, which later influenced soft set similarity research. Majumdar and Samanta [5] introduced the first formal similarity measure specifically designed for soft sets, establishing key mathematical properties and operational frameworks. This was followed by Sulaiman and Mohamad [6], who proposed a Jaccard based similarity measure, and Kharal [7], who developed comprehensive distance and similarity measures for soft sets. Recent advances include specialized similarity measures for extended soft set variants, such as the picture fuzzy soft sets similarity measure by Salsabeela and John [8].

The versatility of soft set theory has led to numerous hybrid approaches that combine soft sets with other mathematical frameworks. Yang et al. [9] explored the combination of interval valued fuzzy sets with soft sets, while Feng et al. [10] investigated the integration of soft sets with both fuzzy and rough sets. More sophisticated extensions include soft rough sets by Zhan et al. [11], and various specialized variants such as ternary fuzzy soft sets [16], cluster soft sets [15], belief interval valued soft sets [13], and confidence soft sets [12]. Hijriati et al. [14] contributed to the theoretical understanding of constructing soft sets from fuzzy subsets, further enriching the theoretical landscape.

Soft set theory has demonstrated remarkable applicability across diverse domains. In healthcare, Gifu [22] showcased innovative applications in healthcare claims analysis, while network analysis applications were explored by Akgüller [21] for social media complexity analysis. Financial applications include stock market analysis by Balcı et al. [19], and decision making frameworks have been extensively developed [17, 18, 20, 26]. Recent innovations include cryptographic applications by Bayram et al. [23] and comprehensive systematic reviews by Alcantud et al. [27], which highlight the growing maturity of the field.

Although soft set applications have flourished in various domains, time series analysis presents a significant research gap. Current approaches to handling uncertainty in temporal data mainly rely on fuzzy logic variants. Johnpaul et al. [30] developed fuzzy representational structures for time series clustering and classification, while Ye et al. [28] focused on global temperature modeling and prediction. Advanced fuzzy approaches include the work by dos Santos Ferreira et al. [33] on fuzzy time series using soft clustering and Santos Ferreira et al. [31] on addressing imprecision in deterministic time series components. Pattanayak et al. [32] contributed fuzzy probabilistic intuitionistic models for forecasting, and recent developments include patient similarity computation for clinical decision support [35]. Contemporary research has witnessed the emergence of sophisticated uncertainty modeling paradigms. Neutrosophic approaches, as demonstrated by Edalatpanah et al. [34]

in financial forecasting and Saqlain et al. [24] in medical diagnosis, offer advanced uncertainty handling capabilities. Pythagorean fuzzy frameworks have gained significant attention, particularly in energy systems evaluation [36] and decision making about rural infrastructure [37]. These advanced approaches, including interval-valued Fermatean neutrosophic systems [25] and multiattribute decision making algorithms [26], represent the current state of the art in uncertainty modeling, but have not yet been systematically integrated with soft set theory for time series applications.

Despite the extensive theoretical development and diverse applications of soft set theory, a critical gap exists in its application to time series analysis. This gap becomes particularly pronounced when considering the inherent characteristics of temporal data that naturally align with the strengths of soft set methodology. Existing time series analysis methods, while mathematically sophisticated, face several fundamental limitations when dealing with real world temporal data. Traditional statistical approaches assume stationary and precise numerical relationships that may not hold in practice [28]. Fuzzy based methods, though more flexible, still require the definition of membership functions, which can be subjective and domain dependent [31, 33]. Even advanced neutrosophic approaches [34], while powerful in handling uncertainty, introduce computational complexity and parameter sensitivity issues that may limit their practical applicability.

Time series data exhibit several characteristics that make it particularly suitable for soft set representation. First, temporal phenomena often involve categorical or qualitative states that cannot be easily quantified using traditional numerical approaches. For instance, climatological patterns may be better described through qualitative descriptors (e.g., "warm periods," "seasonal transitions") rather than precise temperature ranges [29]. Second, time series frequently contain missing or incomplete information, a scenario where soft sets excel due to their parameter-based representation that naturally accommodates partial information.

A fundamental requirement in time series analysis is the ability to measure similarity between different temporal sequences for pattern recognition, classification, and comparative analysis purposes [30, 35]. Current similarity measures in soft set theory [5–7] were primarily designed for static soft sets and may not adequately capture the nuanced relationships present in temporal data. The existing measures may fail to account for the sequential nature of time series or the varying importance of different temporal patterns. Real world applications provide compelling motivations for developing soft-set-based time series analysis methods. In medical diagnosis, particularly heart sound analysis [39], the classification often relies on qualitative patterns that are difficult to quantify precisely but can be naturally represented through soft set parameters. Similarly, climatological analysis [38] involves long term patterns that may be better characterized through flexible, parameter-based representations rather than rigid numerical thresholds.

The convergence of these factors reveals a significant research opportunity: While soft set theory offers theoretical advantages for uncertainty modeling and temporal data naturally exhibits characteristics suitable for soft set representation, no systematic framework exists for integrating these two domains. Furthermore, the lack of specialized similarity measures designed for time series represented as soft sets limits the practical utility of such integration. This gap is particularly notable given the recent advances in related fields.

This study aims to bridge the identified gap between soft set theory and time series analysis through the development of a comprehensive framework that leverages the inherent strengths of soft sets for temporal data modeling. The primary objectives are structured as follows:

Objective 1: To develop a novel ratio based similarity measure specifically designed for soft sets that addresses the limitations of existing similarity measures [5–7] while providing enhanced discriminative capabilities for complex soft set structures.
Objective 2: To establish systematic methodologies for representing time series data within the soft set framework, enabling the natural accommodation of uncertainty, missing values, and qualitative temporal patterns.
Objective 3: To demonstrate the practical effectiveness of the proposed approach through comprehensive real world applications spanning different domains, specifically medical signal processing and climatological data analysis.
Objective 4: To provide theoretical foundations and mathematical characterizations for the proposed methods, ensuring rigorous mathematical grounding and reproducibility.

The remainder of this paper is structured to provide a comprehensive presentation of the theoretical developments, methodological innovations, and practical applications of soft-set-based time series analysis.

The paper is structured as follows: Section 2 establishes the necessary mathematical foundations by reviewing essential concepts from soft set theory [1, 2] and existing similarity measures [5, 7]. Moreover, ratio-based similarity measure for soft sets introduces a novel similarity measure that forms the core theoretical contribution of this work. This section provides formal definitions, mathematical characterizations, and illustrative examples demonstrating the enhanced capabilities compared to existing measures [6, 8].

In Section 3, we present the two proposed methodologies, amplitude threshold soft set and time derivative threshold soft set, for converting time series data into soft set representations. Detailed algorithms are provided to ensure reproducibility.

In Section 4, we demonstrate the practical effectiveness of the proposed approach through comprehensive analysis of the 58-year Brazilian cities temperature dataset [38]. Also, application to medical signal processing validates the approach in the biomedical domain through heart sound classification using PhysioNet data [39]. This section demonstrates the method's effectiveness in distinguishing between normal and abnormal heart sounds, contributing to soft set applications in healthcare [22].

In Section 5, we address the comparative analysis gap identified in current literature while highlighting the complementary nature of the approach relative to recent advances.

In Section 6, we summarize the key contributions and outlines promising avenues for future research, including extensions to multivariate time series and integration with emerging uncertainty modeling paradigms.

Figure 1 shows the overall research flow adopted in this study. This visual summary provides a clear road map of how the research progresses from the initial stage to the conclusion.

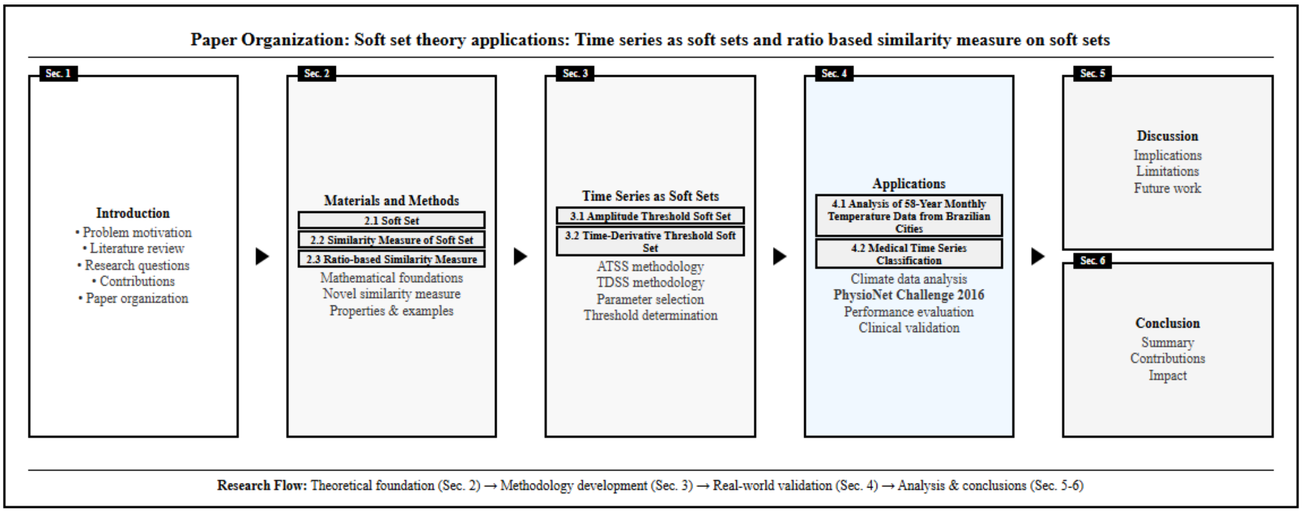

Figure 1. Research flow.

2. Materials and methods

2.1. *Soft set*

The notation U, $P(U)$, and E stand for the initial universe, power set of the universe, and the set of parameters, respectively.

Definition 1. *[1] A soft set (F, E) over U is defined by a set-valued function $F : E \longrightarrow P(U)$. Soft set (F, E) can be represented by a set of ordered pairs such that $(F, E) = \{(x, F(x)) : x \in E, F(x) \in P(U)\}$.*

Definition 2. *[19] Let (F, A) be a soft set on the initial universe U. For $u \in U$, the number of $F(e)$'s that include u is called the soft degree of u, where $e \in A$ and is denoted by $\delta(u)$. Similarly, the cardinality $\mid F(e_i) \mid$ is called the soft degree of e_i.*

2.2. *Similarity measure of soft set*

The similarity measure between two soft sets has been defined by numerous authors. Various techniques exist for defining such measures; some are based on distance metrics, while others rely on matching functions. In addition, there are methods rooted in set-theoretic approaches. Certain properties are shared across these measures, while others differ, influencing the selection of an appropriate measure for different applications. In this section, a new measure for soft sets will be defined as an extension of the Jaccard-based similarity measure for soft sets, which is defined by [6] as follows.

Definition 3. *[6] Let $\{U\} = \{x_1, x_2, \dots, x_n\}$ be a universe of elements, and E be a set of parameters. Suppose (F, A) and (G, B) are two soft sets in the soft class (U, E) such that $A, B \subset E$. We define a Jaccard-based set theoretic similarity measure S_J as*

$$S_J[(F, A), (G, B)] = \begin{cases} \frac{\sum_{e \in A \cap B} \omega_e \cdot |F(e) \cap G(e)|}{|F(e) \cup G(e)|} & \text{if } A \cap B \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

$|A|$ refers to the cardinality of the set A, and, for all $e \in A \cap B$, ω_e is the weight proportion of the parameter e such that $0 < \omega_e < 1$ (overlapping parameter(s) have nonzero weight).

Ratio-based similarity measure

The extended similarity measure model generalizes several set-theoretical similarity models proposed in the literature. Let (F, A) and (G, B) be soft sets over the common universe U. The similarity between these soft sets is defined as

$$S_R((F,A),(G,B)) = \frac{1}{\mid A \cup B \mid}.\sum_{e \in A \cap B} \frac{\mid F(e) \cap G(e) \mid}{\mid F(e) \cap G(e) \mid + \alpha \mid F(e) \cap G(e)^c \mid + \beta. \mid F(e)^c \cap G(e) \mid}.$$

In this formula, two alternative choices can be considered for α and β.

Case 1. (Raw degree-based weights) α, β soft degree of the parameter e with respect to soft sets (F, A) and (G, B), respectively. Raw cardinalities are used when each parameter e is assigned weights directly based on the number of elements in the associated approximate set:

$$\alpha = |F(e)|, \quad \beta = |G(e)|.$$

This approach penalizes mismatches more heavily, as each unmatched element is scaled by the full size of its respective soft set component. As such, even small differences between soft sets can result in significantly reduced similarity values.

This method is suitable for scenarios where:

- The number of parameters is relatively small, and each parameter carries significant decision weight.
- Minor discrepancies are critical, such as in medical diagnostics, anomaly detection, or security systems.
- Parameter sizes reflect their impact or importance explicitly.

Case 2. (Normalized weight) For $F(e) \cap G(e) \neq \emptyset$,

$$\alpha = \frac{|F(e)|}{\sum_{e \in A} |F(e)|}, \quad \beta = \frac{|G(e)|}{\sum_{e \in B} |G(e)|},$$

where $\alpha, \beta \geq 0$ are adjustable parameters to fine-tune the influence of different types of mismatch.
In contrast, normalized weighting uses the proportion of elements with respect to the entire universe U:

$$\alpha = \frac{|F(e)|}{\sum_{e \in A} |F(e)|}, \quad \beta = \frac{|G(e)|}{\sum_{e \in B} |G(e)|}.$$

This strategy softens the penalty for mismatches and allows the similarity measure to reflect broader pattern matching rather than strict element-wise agreement. It is particularly effective when dealing with large universes or high-dimensional soft sets with many parameters.

This approach is recommended when:

- The universe size is large, and parameter sets are relatively sparse or variable.
- General pattern similarity is more important than exact matches, such as in recommendation systems or clustering.
- Fairness and balance are desired when comparing heterogeneous soft sets.

Summary of use cases for raw degree-based weights and normalized weights are given in the following Table 1.

Table 1. Comparison of usage of similarity measures.

Condition	Preferred weighting strategy
Few parameters with critical significance	Raw degree-based
High sensitivity to mismatch required	Raw degree-based
Large universe with many parameters	Normalized weights
Pattern-oriented similarity needed	Normalized weights
Similarity in recommendation or ML systems	Normalized weights

Throughout this paper, raw degree-based weights are used to measure similarity between soft sets. The comparison of different similarity measures defined on soft sets was stated in the following example by Sulaiman and Mohamad in [6]. The same example is reconsidered with this new similarity measure called the ratio-based similarity measure, and the results are shown by adding a new column to the table given earlier by Sulaiman and Mohamad in [6]. In the ratio-based similarity measure for soft sets, the choice of parameter weights plays a critical role in determining the behavior and interpretation of similarity. Two common strategies for assigning these weights are: Using raw cardinalities (raw degrees), and normalized values with respect to the universe. The appropriateness of each strategy varies with the context and structure of the data.

Example 1. *[6] Let $U = \{p, q, r, s\}$ and $E = \{e_1, e_2, e_3\}$ be the universe set of objects and the universe set of parameters, respectively. Consider the following soft sets in a soft space (U, E):*

- $(K, D) = \{e_1 = \{p, q\}, e_2 = \{q, s\}\}$,
- $(L, M) = \{e_1 = \{p, q\}, e_2 = \{q, s\}, e_3 = \{r, s\}\}$,
- $(H, C) = \{e_1 = \{p\}, e_2 = \{q, s\}\}$,
- $(R, J) = \{e_1 = \{p, q\}, e_2 = \{q, s\}, e_3 = \{p, r, s\}\}$,
- $(G, B) = \{e_2 = \{q, s\}\}$,
- $(P, N) = \{e_1 = \{p\}, e_2 = \{q\}, e_3 = \{r\}\}$,
- $(T, W) = \{e_1 = \{p, q\}, e_2 = \{q, r, s\}\}$.

In Table 2 similarity values by different similarity measures for defined soft sets. The last column contains ratio-based similarity measures of soft sets.

Table 2. Similarity values by different similarity measures for soft sets.

Pair of soft sets	S_1	S_2	S_3	S_4	S_5	S_J	S_R
$(K, D), (L, M)$	0.55	0.67	0.67	0.50	0.57	0.67	0.66
$(G, B), (P, N)$	0.50	0.50	0.25	0.25	0.35	0.17	0.06
$(P, N), (T, W)$	0.39	0.43	0.33	0.24	0.34	0.28	0.16
$(R, J), (L, M)$	0.63	0.75	0.88	0.50	0.63	0.89	0.8
$(L, M), (P, N)$	0.50	0.50	0.50	0.37	0.45	0.50	0.33

The similarity measures symbolized by S_1, S_2, S_3, S_4, S_5, and S_6 were stated in [6]. Looking at *the results in this table, it can be observed that S_R works as a more stringent or sensitive similarity measure. If selectivity or small differences in similarity values are desired, S_R may be a suitable choice.*

Example 2. *Consider the soft sets (F, A) and (G, B) defined on the parameter set and universal set, given as $E = \{a_1, a_2, a_3\}$ and $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$, where $E = A$ and $B = \{a_2\}$ with the set valued functions*
$F(a_1) = \{x_1, x_4, x_6\}$,
$F(a_2) = \{x_1, x_2, x_3\}$,
$F(a_3) = \{x_1, x_5, x_7\}$, *and*
$G(a_2) = \{x_2, x_4, x_7\}$.
$A \cap B = \{a_2\}$. *The degree of parameter a_2 with respect to both soft sets is 3. The similarity of these soft sets is* 0.026. $S_R((F, A), (G, B)) = 0.026$.

Proposition 1. *Let (F_1, E_1), (F_2, E_2), and (F_3, E_3) be three soft sets over the same finite universe U. Then, the following hold:*

1. $S_R((F_1, E_1), (F_2, E_2)) = S((F_2, E_2), (F_1, E_1))$,
2. $0 \leq S_R((F_1, E_1), (F_2, E_2)) \leq 1$,
3. $S_R((F_1, E_1), (F_1, E_1)) = 1$.
4. *If* $(F_1, E_1) \subseteq (F_2, E_2)$ *and* $(F_2, E_2) \subseteq (F_3, E_3)$, *then* $S_R[(F_1, E_1), (F_3, E_3)] \leq S_R[(F_1, E_1), (F_2, E_2)]$ *and* $S_R[(F_1, E_1), (F_3, E_3)] \leq S_R[(F_2, E_2), (F_3, E_3)]$.

Proof. 1. Observe that $E_1 \cup E_2 = E_2 \cup E_1$ and $E_1 \cap E_2 = E_2 \cap E_1$. For each $e \in E_1 \cap E_2$, the term $|F_1(e) \cap F_2(e)|$ is symmetric in F_1 and F_2. The denominators are also symmetric because:

$$|F_1(e) \cap F_2(e)^c| = |F_2(e)^c \cap F_1(e)|, \text{ and } |F_1(e)^c \cap F_2(e)| = |F_2(e) \cap F_1(e)^c|,$$

and the soft degrees α and β are assigned to their respective soft sets. Thus, the entire expression is symmetric.

2. For each $e \in E_1 \cap E_2$, the term

$$\frac{|F_1(e) \cap F_2(e)|}{|F_1(e) \cap F_2(e)| + \alpha|F_1(e) \cap F_2(e)^c| + \beta|F_1(e)^c \cap F_2(e)|}$$

is a ratio where the numerator is a subset of the denominator. Thus, each term is in $[0, 1]$. The sum of such terms divided by $|E_1 \cup E_2|$ is also in $[0, 1]$.

3. When $(F_1, E_1) = (F_2, E_2)$, we have $E_1 = E_2$, so $E_1 \cup E_2 = E_1$ and $E_1 \cap E_2 = E_1$. For each $e \in E_1$, the term becomes

$$\frac{|F_1(e) \cap F_1(e)|}{|F_1(e) \cap F_1(e)| + \alpha|F_1(e) \cap F_1(e)^c| + \beta|F_1(e)^c \cap F_1(e)|} = \frac{|F_1(e)|}{|F_1(e)|} = 1.$$

The sum over $e \in E_1$ is $|E_1|$, and dividing by $|E_1|$ gives 1.

4. The inclusion $(F_1, E_1) \subseteq (F_2, E_2)$ implies that for all $e \in E_1 \cap E_2$, $F_1(e) \subseteq F_2(e)$. Similarly, $(F_2, E_2) \subseteq (F_3, E_3)$ implies $F_2(e) \subseteq F_3(e)$. For $e \in E_1 \cap E_3$, we have $F_1(e) \subseteq F_2(e) \subseteq F_3(e)$, so

$$\frac{|F_1(e) \cap F_3(e)|}{|F_1(e) \cap F_3(e)| + \beta|F_3(e) \setminus F_1(e)|} \leq \frac{|F_1(e) \cap F_2(e)|}{|F_1(e) \cap F_2(e)| + \beta|F_2(e) \setminus F_1(e)|},$$

since $|F_3(e) \setminus F_1(e)| \geq |F_2(e) \setminus F_1(e)|$. Summing over all such e preserves the inequality. The second part follows similarly.

□

Incorporating α and β as degrees of parameters in the ratio-based similarity measure adds a layer of contextual sensitivity that is missing in the standard Jaccard similarity. It enables the measure to dynamically adapt to the importance of parameters, provide more meaningful and nuanced similarity scores, and to better handle mismatches based on the significance of the parameters. The example presented in [7] is re-evaluated using the proposed method, employing the ratio-based similarity measure instead.

Example 3. *Let us consider the financial diagnostic problem given in [7]. The universe X consists of critical financial metrics:*

$$X = \begin{Bmatrix} s: \textit{share price}, & p: \textit{profit-earning ratio}, \\ i: \textit{inflation}, & c: \textit{competitiveness}, \\ o: \textit{future outlook}, & m: \textit{paid up capital}, \\ d: \textit{business diversification}, & f: \textit{foreign direct investment}, \\ l: \textit{debt level}, & x: \textit{fixed income} \end{Bmatrix}.$$

The parameter set E represents qualitative states: $E = \begin{Bmatrix} e_1: \textit{high}, & e_2: \textit{rising}, \\ e_3: \textit{low}, & e_4: \textit{fluctuate}, \\ e_5: \textit{bearish} & \end{Bmatrix}.$

The model soft set (H, C) *is determined as* $H(e_1) = \{s, o\}, H(e_2) = \{c\}, H(e_4) = \{i, l\}, H(e_5) = \{p, f\}$.

Profiles of firms ABC and XYZ under study are given with the soft sets (F, A) *and* (G, B)*, respectively, as follows*

$(F, A) = \{(e_2, \{s, f\}), (e_3, \{p, i\}), (e_5, \{o, s\})\}$,

$(G, B) = \{(e_1, \{s, o\}), (e_2, \{c\}), (e_3, \{m, i\}), (e_4, \{i, l\}), (e_5, \{p, f\})\}$.

We employ the ratio-based similarity measure to find the similarity between the firms and the model soft set:

Similarity: $S_R((H, C), (F, A)) = 0, S_R((H, C), (G, B)) = 0.8$. *On the other hand, similarity was estimated as* $S((H, C), (F, A)) = 0.15, S((H, C), (G, B)) = 0.5$ *by [7].*

Inspired by this example, it can be stated that the similarity measure introduced in our study is more practical. Moreover, the similarity measure defined for soft sets can also be utilized in data analysis. To this end, two different methods will first be presented to represent time series using soft sets. Then, using the similarity measure defined for soft sets, it is demonstrated that interpretations can be made for given datasets.

3. Time series as soft sets

Soft set theory, introduced by Molodtsov in 1999, provides a mathematical framework for handling uncertainties and vagueness inherent in complex systems. When applied to time-series data, soft sets offer a flexible approach to capture and analyze underlying patterns, relationships, and dynamics that may not be evident through traditional methods. By defining appropriate criteria or thresholds based on statistical properties of the time series, we can construct soft sets that encapsulate various aspects such as amplitude levels or rates of change. This methodology allows for a nuanced examination of time-series data, revealing intricate structures within the data.

The soft set is represented as a pair (F, E), where:
- E is the criterion based on a specific property of the time series.
- F is the set of observation values at time points that satisfy the condition imposed by E.

By adjusting the parameters within the criteria, the soft sets can be tailored to focus on particular features of the time series. This flexibility allows us to capture subtle patterns and behaviors that may be critical for understanding the system's dynamics.

Various methods are introduced in this article. As in the first method, soft sets are derived from the peaks and anomaly values of individual time series. In the second method, time-derived soft sets are defined, which provide a new approach to represent time series, capturing sudden changes and anomalies through dynamic behaviors such as rates of change and amplitude variations. In addition, time-derived soft sets were examined with the newly-developed Jaccard-based similarity method to measure the similarity between different time series. A data set with 58 years of monthly temperature values was selected to apply these new methods as an application area.

3.1. Amplitude threshold soft set

The amplitude threshold soft set focuses on identifying observation values in a time series that are similar within a specified threshold. Consider a time series $X = \{x_t \mid t = 1, 2, \ldots, N\}$, where x_t represents the amplitude at time t.

The criterion E_{amp} is defined based on the standard deviation $\sigma(x)$ of the amplitude values. The amplitude threshold is set as

$$E_{\text{amp}} = \alpha \cdot \sigma(x),$$

where α is a scaling factor (e.g., $\alpha = 0.1$) that determines the sensitivity of the threshold.

We select a reference value x_{ref}, typically the mean $\mu(x)$ of the time series,

$$x_{\text{ref}} = \mu(x) = \frac{1}{N} \sum_{t=1}^{N} x_t.$$

The soft set elements F_{amp} consist of the observation values x_t at time points where the absolute difference from the reference value is less than E_{amp}:

$$F_{\text{amp}} = \left\{ x_t \mid |x_t - x_{\text{ref}}| < E_{\text{amp}} \right\}.$$

Thus, the soft set is represented as $(F_{\text{amp}}, E_{\text{amp}})$, capturing all observation values that are similar to the reference amplitude within the specified threshold. This method highlights periods where the amplitude remains within a certain range, which can be crucial for identifying steady states or consistent patterns in the time series.

3.2. Time-derivative threshold soft set

The time-derivative threshold soft set examines the observation values of the rate of change in the time series, providing insight into its dynamic behavior. The first differences Δx_t are calculated to

approximate the time derivatives:
$\Delta x_t = x_{t+1} - x_t, \quad \text{for } t = 1, 2, \ldots, N-1.$
The standard deviation $\sigma(\Delta x)$ of these differences quantifies the variability in the rate of change. The criterion F_{der} is then defined as

$$\sigma = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\Delta(X_t) - \mu)^2}.$$

The standard deviation $\sigma(\Delta x)$ of these differences quantifies the variability in the rate of change. The criterion F_{der} is then defined as

$$F_{\text{der}} = \beta \cdot \sigma(\Delta x).$$

A reference is selected as the rate of change Δx_{ref}, which could be the mean of the first differences:

$$\Delta x_{\text{ref}} = \mu(\Delta x) = \frac{1}{N-1}\sum_{t=1}^{N-1}\Delta x_t.$$

The soft set elements F_{der} include the observation values Δx_t, where the absolute difference from the reference rate is less than E_{der}:

$$F_{\text{der}} = \{x_t \mid |\Delta x_t - \Delta x_{\text{ref}}| < E_{\text{der}}\}.$$

The soft set $(F_{\text{der}}, E_{\text{der}})$ captures rates of change that are similar to the average rate within the specified threshold. This approach is useful for identifying periods where the time series exhibits consistent dynamics, such as steady growth or decline.

4. Aplications

Example 4. *This example applies the the amplitude threshold soft set theory to analyze daily temperature patterns in Istanbul and Ankara during January 2023. We demonstrate how this method identifies normal temperature ranges and detects anomalies in urban climate data. Table 3 presents the complete January 2023 temperature data and Table 4 contains some statistical properties of Ankara and İstanbul:*

Table 3. Daily temperature records for Istanbul and Ankara (January 2023).

Date	Istanbul (°C)	Ankara (°C)
2023-01-01	8.5	2.1
2023-01-02	7.2	0.8
2023-01-03	6.8	-1.2
2023-01-04	5.1	1.5
2023-01-05	9.3	3.4
2023-01-06	10.2	5.6
2023-01-07	11.5	4.2
2023-01-08	12.0	0.7
2023-01-09	4.9	-2.1
2023-01-10	3.7	-2.3
...	...	...
2023-01-31	6.2	-1.4

Table 4. Statistical comparison.

Metric	Istanbul	Ankara
Mean (μ)	8.2 °C	1.8 °C
Std. Dev. (σ)	2.3	3.1
Threshold (E_{amp})	1.2	1.6
Normal Range	[7.0, 9.4]	[0.2, 3.4]

Istanbul temperature soft set for Istanbul's daily temperatures with $\mu_I = 8.2$ °C *and* $\sigma_I = 2.3$ °C*:*
$E^I_{amp} = 0.5 \times \sigma_I = 1.15$ °C,
$F^I_{amp} = \{8.5, 7.2, 9.3, 7.8, 8.1, 9.2, 7.5, 6.3, 8.4, 7.1, 6.7, 7.2, 8.9, 9.1, 8.7, 7.4, 6.9, 6.5, 6.2\}$,
$F^I_{amp} = \{x_1, x_2, x_5, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{18}, x_{19}, x_{20}, x_{24}, x_{25}, x_{26}, x_{27}, x_{28}, x_{29}, x_{30}, x_{31}\}$.
Ankara temperature soft set for Ankara's daily temperatures with $\mu_A = 1.8$ °C *and* $\sigma_A = 3.1$ °C*:*
$E^A_{amp} = 0.5 \times \sigma_A = 1.55$ °C,
$F^A_{amp} = \{2.1, 0.8, 1.5, 3.4, 0.7, 1.2, 0.5, 2.8, 0.5, 1.1, 0.7, 1.5, 0.9\}$,
$F^A_{amp} = \{x_1, x_2, x_4, x_5, x_7, x_{11}, x_{12}, x_{13}, x_{15}, x_{18}, x_{25}, x_{26}, x_{27}\}$.
We know check the ratio-based similarity of obtained soft sets $S_R((F^A_{amp}, E^A_{amp}), (F^I_{amp}, E^I_{amp})) = 0.0582$. *The 5.8 % similarity coefficient quantitatively confirms the pronounced microclimatic differentiation between the coastal (Istanbul) and continental (Ankara) urban environments.*

Example 5. *In the following tables, Tables 5 and 6, we present two time series tracking the symptoms of COVID-19 in a patient and healthy person over the course of one week. The variables being monitored are body temperature, oxygen saturation, and cough severity,*

Table 5. 7-day vital signs of a sick individual.

Day	a_1=Body temperature (°C)	a_2=Oxygen saturation (%)	a_3=Cough severity (1-10)
1	37.8	95	4
2	38.5	93	6
3	39.1	90	8
4	38.9	88	7
5	38.2	92	5
6	37.5	94	4
7	37.2	96	3

Table 6. Vital signs of a healthy individual over 7 days.

Day	a_1=Body temperature (°C)	a_2=Oxygen saturation (%)	a_3=Cough severity (0-5)
1	36.7	98	0
2	36.5	97	0
3	36.7	98	0
4	36.6	99	0
5	36.7	98	0
6	36.6	97	0
7	36.8	98	0

There are three different time series in this example, including body temperature: $= [37.8, 38.5, 39.1, 38, 9, 38.2, 37.5, 37.2]$, *oxygen saturation*:$= [95, 93, 90, 88, 92, 94, 96]$ *and cough severity (1-10)*:$= [4, 6, 8, 7, 5, 4, 3]$. *By using the mentioned methods, new soft sets based on these data are constructed.*

Next we construct four soft sets (F_{amp_1}, E), (F_{der_1}, E), (F_{amp_2}, E), *and* (F_{der_2}, E) *according to this time series by using the methods introduced in the previous section. We have the set of parameters* $E = \{a_1, a_2, a_3\}$, *where the parameters stand for* a_1*=body temperature (°C),* a_2*=oxygen saturation and* a_3*=cough severity respectively.*

Amplitude threshold soft set: *The mean values and standard deviation of each series are, respectively, body temperature (°C): 38.17, oxygen saturation: 92.57 and cough severity (1–10): 5.29: and body temperature (°C): 0.658, oxygen saturation: 2.611, and cough severity (1–10): 1.666. The threshold* E_{amp} *is chosen as the standard deviation.*

$F_{amp_1}(a_1) = \{x_1, x_2, x_5\}$,
$F_{amp_1}(a_2) = \{x_1, x_2, x_3, x_5, x_6\}$,
$F_{amp_1}(a_3) = \{x_1, x_2, x_5, x_6\}$.
$F_{amp_2}(a_1) = \{x_1, x_3, x_4, x_5, x_6\}$,
$F_{amp_2}(a_2) = \{x_1, x_3, x_5, x_7\}$,
$F_{amp_2}(a_3) = \{0\}$.
$S_R(F_{amp_1}, F_{amp_2}) = \frac{15}{85}$.

Time-derivative threshold soft set: *The mean values and the standard deviations are, respectively,* -0.1, *0.17, and* -0.17; *and 0.57, 2.61, and 1.57 for time derivative threshold soft sets.* $E_{der} = \sigma(\Delta x)$,

$F_{der_1}(a_1) = \{x_4, x_6\}$,
$F_{der_1}(a_2) = \{x_1, x_3, x_5, x_6\}$,
$F_{der_1}(a_3) = \{x_3, x_5, x_6\}$.
The second soft set (F_2, E) *is constructed as*
$F_{der_2}(a_1) = \{x_3, x_4, x_5\}$,
$F_{der_2}(a_2) = \{\}$,
$Fder_2(a_3) = \{\}$.
The calculated daily differences in cough severity is zero for each term. The calculated standard deviations (σ) for the daily differences are as follows:

- *Body temperature:* $\sigma = 0.158$.
- *Oxygen saturation:* $\sigma = 1.0$.

The similarity between these two soft sets is 0.33, which is less than $\frac{1}{2}$*. The analysis of the values obtained has further strengthened the conclusion that there exists a notable distinction between the soft set derived from the vital parameters of a healthy individual and the soft set generated from the vital signs of a patient. This divergence highlights how specific health indicators, when observed over time, can reflect underlying differences between health statuses, thus providing a clearer, more structured comparison between the two sets of data*

4.1. Analysis of 58-year monthly temperature data from brazilian cities using soft sets

Example 6. *In this example, soft sets will be constructed from a dataset that contains monthly temperature measurements from Brazilian cities from 1967 to 2019. This data set contains 58-year temperatures of Brazilian cities. The 58 years of monthly temperature data for six cities represent time series for each city. Using the method described above, soft sets were created for each city. The parameter set and the initial universe are the same for all soft sets. The parameter set consists of the months, while the universe consists of the years. From the time series of these six cities, six soft sets were constructed for the Brazilian cities Salvador, Goiania, Fortaleza, Curitiba, Belem, and Vitoria. Map of the this 6 cities is given in Figure 2. This data, obtained from Kaggle, was evaluated using soft sets, and with this representation, some analysis can be performed on this data. Before applying the introduced methods, some data statistics are given to understand the data sets better.*

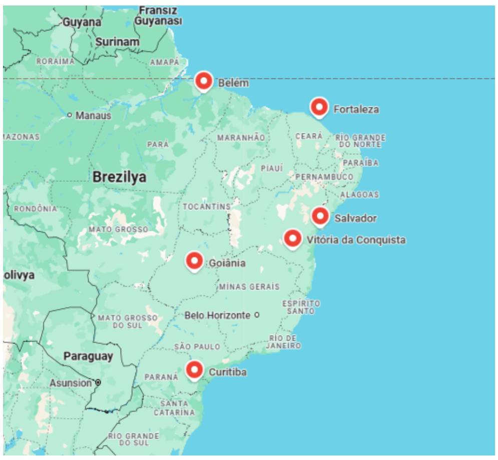

Figure 2. Map of the cities considered in the example.

There are many advantages to representing a given time series with soft sets. The methods amplitude threshold soft set and time derivative threshold soft set will be applied to this dataset. After building soft sets for each city, the similarities between these soft sets will be analyzed. As a result of this analysis, it will be observed which cities experienced similar temperature changes over the 58-year period. The parameter sets of the soft sets are the months, and the universal sets are the years.

The 6 different soft sets (F_b, E), (F_c, E), (F_g, E), (F_s, E), (F_v, E), *and* (F_f, E) *are obtained for Belem, Curitiba, Goiania, Salvador, Vitoria, and Fortaleza with these two methods.*

This dataset contains 58 years of monthly temperature values for six cities. The data have been evaluated separately for each city, both by visually representing the Jaccard similarity matrices derived from binary matrices created using the matrix representation provided by soft sets, and by analyzing the datasets individually. Finally, the temperature changes observed in these cities over the years will be compared by applying the similarity measures defined on soft sets.

As a first step, the Jaccard similarity matrix is obtained for the matrix representation of each soft set (F_b, E), (F_c, E), (F_g, E), (F_s, E), (F_v, E), *and* (F_f, E). *Annual and monthly average temperatures for each city are presented in the following figures. The Jaccard similarity matrices are visualized using the hierarchical clustering dendrograms shown below. Figures 3 and 4 show annual and monthly average temperatures of six cities, respectively. Figures 5 and 6 show hierarchical clustering dendrogram of six cities, respectively..*

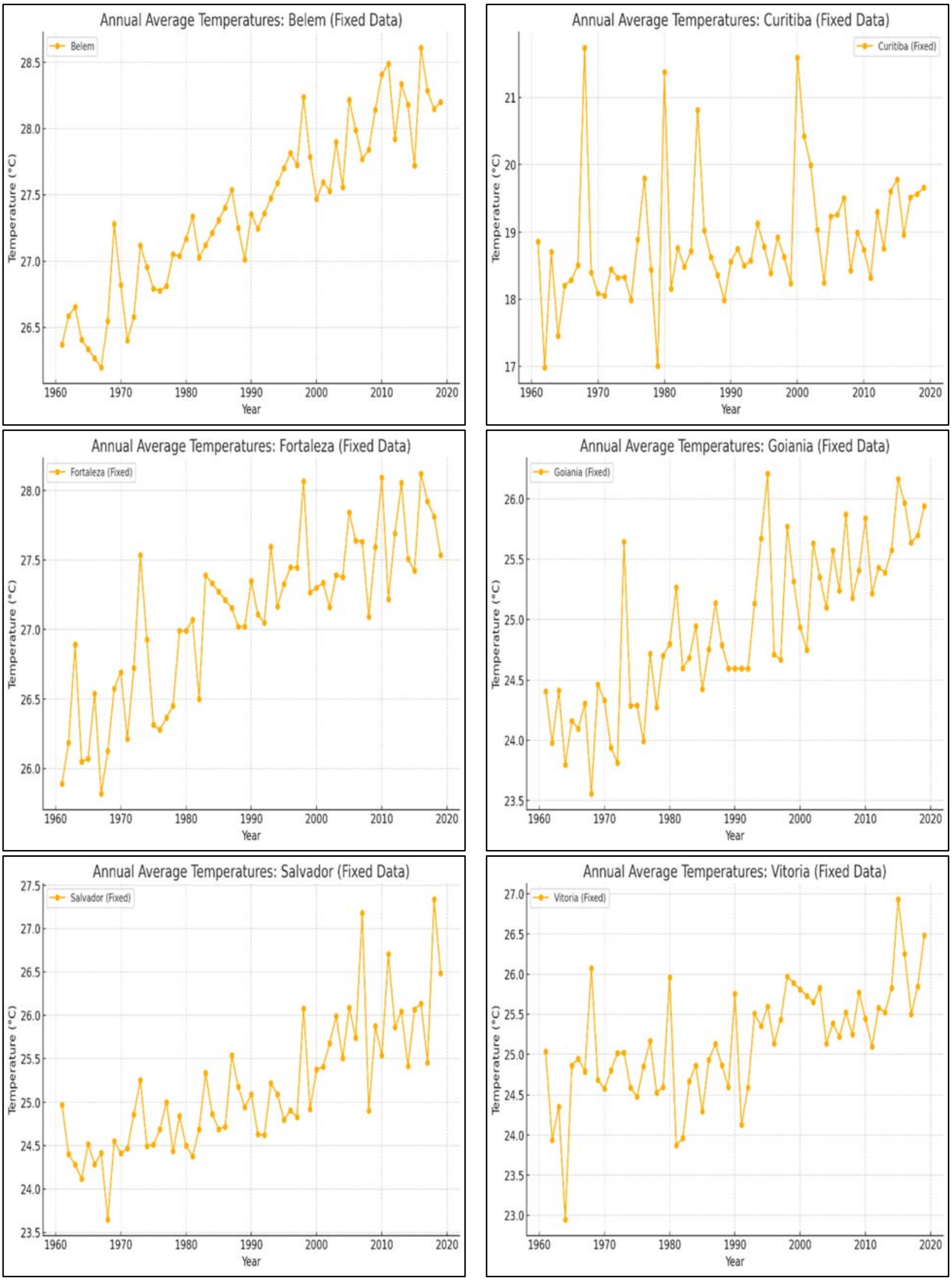

Figure 3. Annual average temperatures of six cities (fixed data).

Figure 4. Annual average temperatures of six cities (fixed data).

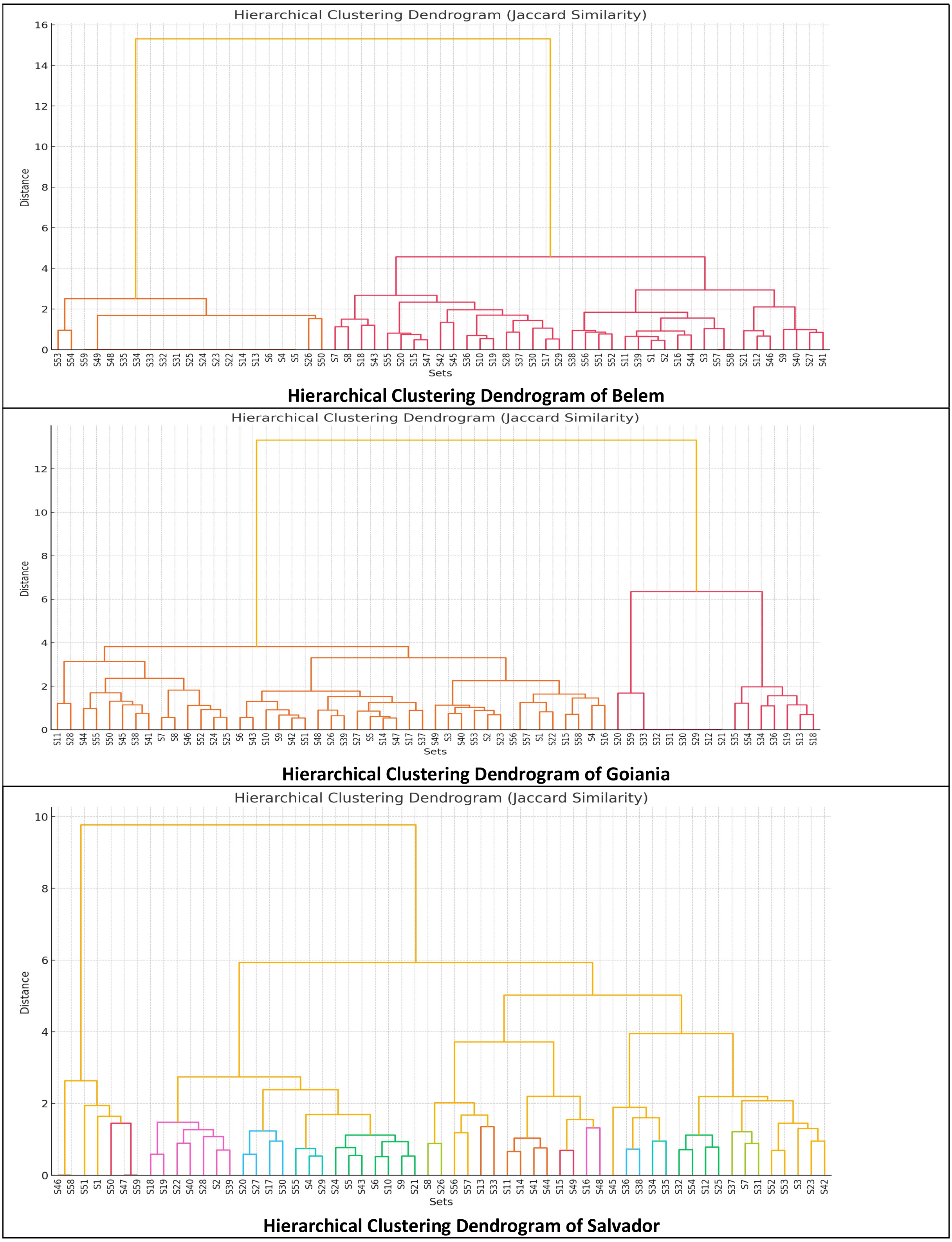

Figure 5. Hierarchical clustering dendrogram.

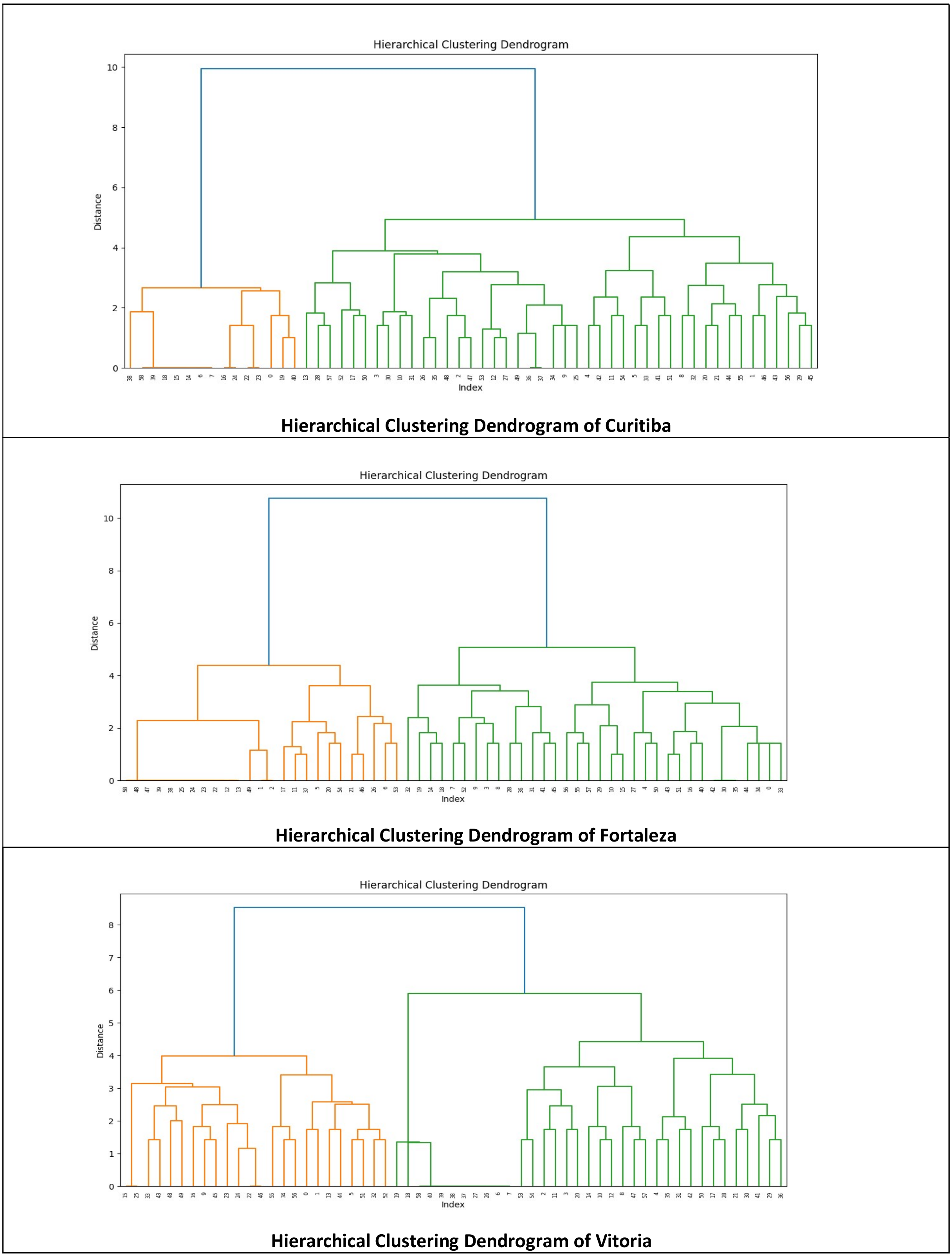

Figure 6. Hierarchical clustering dendrogram.

Pairwise similarities of cities are evaluated by using the ratio-based similarity measures of soft sets and listed in Table 7.

Table 7. Similarity values of cities.

Pair	Similarities of T.D. threshold soft set	A. threshold soft set
$S_R(F_b, F_c)$	0.16	0.34
$S_R(F_b, F_f)$	0.24	0.69
$S_R(F_b, F_g)$	0.23	0.40
$S_R(F_b, F_s)$	0.17	0.39
$S_R(F_b, F_v)$	0.15	0.35
$S_R(F_c, F_f)$	0.23	0.31
$S_R(F_c, F_g)$	0.21	0.37
$S_R(F_g, F_f)$	0.17	0.43
$S_R(F_g, F_s)$	0.26	0.46
$S_R(F_g, F_v)$	0.29	0.48
$S_R(F_s, F_c)$	0.24	0.32
$S_R(F_s, F_f)$	0.20	0.45
$S_R(F_s, F_g)$	0.26	0.46
$S_R(F_s, F_v)$	0.23	0.45
$S_R(F_v, F_c)$	0.29	0.45
$S_R(F_v, F_f)$	0.23	0.35
$S_R(F_v, F_g)$	0.29	0.48

Figure 7 provides a comprehensive visualization of all similarity values of soft sets.

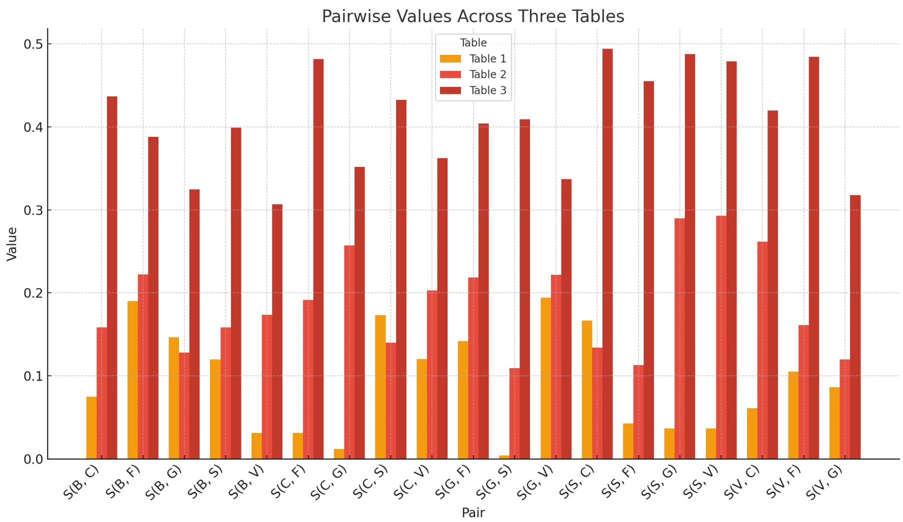

Figure 7. Overall process.

4.2. *Medical time series classification*

This subsection demonstrates the application of the proposed soft set methodology to medical time series analysis using the PhysioNet Challenge 2016 dataset for heart sound classification. This application directly addresses the clinical relevance of soft set theory in handling uncertainty and noise inherent in medical signals. Figure 8 presents the main steps of the proposed framework, from preprocessing of medical time series data to applying the cluster-based classification method and the final classification process.

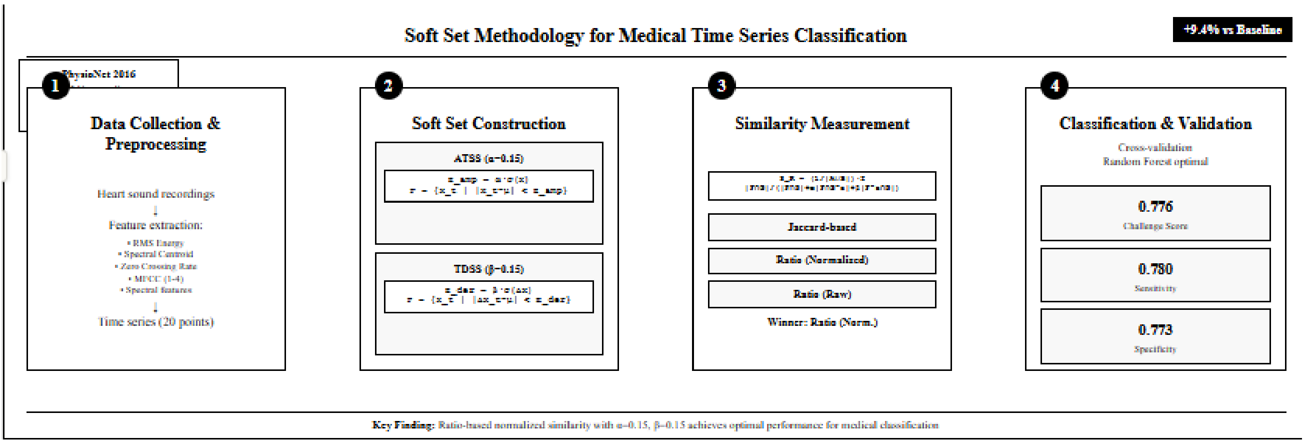

Figure 8. Methodology flowchart for soft-set-based medical time series classification.

4.2.1. Dataset and problem description

The PhysioNet Computing in Cardiology Challenge 2016 focused on heart sound segmentation and classification, providing a standardized benchmark for evaluating automated cardiac diagnostic systems. The dataset comprises 3,541 heart sound recordings collected from multiple clinical sites, with recordings varying in length from 5 seconds to over 120 seconds.

The classification task involves distinguishing between:

- Normal heart sounds (Class 0): Healthy cardiac activity.
- Abnormal heart sounds (Class 1): Pathological conditions including murmurs, extra heart sounds, and irregular rhythms.

4.2.2. Heart sound feature extraction and soft set construction

Heart sound recordings were processed using spectral and temporal feature extraction techniques specifically adapted for cardiac signals. The following features were extracted with a target length of 20 time points per recording:

1. RMS energy: Root mean square energy indicating signal amplitude variations.
2. Spectral centroid: Frequency distribution center, relevant for heart sound characterization.
3. Zero crossing rate: Temporal dynamics indicator.
4. Spectral rolloff: High-frequency content measure.
5. MFCC coefficients: Mel-frequency cepstral coefficients (first 4 components).
6. Spectral bandwidth: Frequency spread measure.
7. Chroma features: Harmonic content representation.

For each extracted feature time series, both amplitude threshold soft sets (ATSS) and time derivative threshold soft sets (TDSS) were constructed using the methodology defined in Section 3. The parameter values α and β control the sensitivity of threshold determination, with optimal values determined through systematic evaluation.

4.2.3. Comprehensive similarity measure evaluation

A critical contribution of this study is the systematic comparison of similarity measures on medical data. Three similarity approaches were evaluated using Jaccard-based similarity, ratio-based similarity (normalized) and ratio-based similarity (raw). The definition of the similarity measures are given in Section 2.

4.2.4. Experimental design and validation

The experimental evaluation followed rigorous medical validation standards:

Dataset split: Subject-based stratified split with 70% training (700 recordings) and 30% testing (300 recordings) to prevent data leakage across subjects.

Cross-validation: 5-fold stratified group K-fold cross-validation ensuring subjects appear in only one fold.

Classification models: Three established algorithms were employed:

- Random forest (n_estimators=200, max_depth=10).
- Gradient boosting (n_estimators=100, max_depth=6).
- Logistic regression with L2 regularization.

Evaluation metrics: PhysioNet challenge 2016 official metrics:

- Challenge score: $\frac{\text{Sensitivity+Specificity}}{2}$.
- Sensitivity: True positive rate.
- Specificity: True negative rate.
- Additional metrics: F1-score, AUC, accuracy.

4.2.5. Results and performance analysis

Parameter optimization results. Systematic evaluation across parameter ranges ($\alpha, \beta \in \{0.1, 0.15, 0.2, 0.25, 0.3\}$) revealed optimal performance at $\alpha = 0.15$, $\beta = 0.15$. The effect of different values of α and β can be observed from Table 8.

Table 8. Parameter sensitivity analysis.

Parameters	Challenge score	F1-score	AUC
$\alpha = 0.1, \beta = 0.1$	0.764	0.764	0.797
$\alpha = 0.15, \beta = 0.15$	**0.776**	**0.777**	0.798
$\alpha = 0.2, \beta = 0.2$	0.770	0.770	**0.809**
$\alpha = 0.25, \beta = 0.25$	0.775	0.774	0.806
$\alpha = 0.3, \beta = 0.3$	0.771	0.770	0.812

Similarity measure comparison. The ratio-based normalized similarity measure demonstrated superior performance for medical time series classification:

Best performance results ($\alpha = 0.15, \beta = 0.15$):

- Method: Ratio-based normalized similarity.
- Classifier: Random forest.
- Challenge score: 0.776 ± 0.028.
- Sensitivity: 0.780.
- Specificity: 0.773.
- F1-score: 0.777
- AUC: 0.798.

Table 9 summarizes the benchmark comparison with the PhysioNet challenge results. It reports the performance of the proposed method alongside baseline and reference models from the PhysioNet challenge.

Table 9. Benchmark comparison with PhysioNet challenge results.

Method	Challenge score	Sensitivity	Specificity
PhysioNet Winner (2016)	0.86	0.94	0.78
PhysioNet Baseline	0.71	0.65	0.76
Our Soft Set Method	**0.776**	**0.780**	**0.773**
Improvement over Baseline	**+9.3%**	**+20.0%**	**+1.7%**

4.2.6. Clinical and methodological implications

Why soft sets for medical data? The superior performance of soft set methodology in heart sound classification can be attributed to several factors:

1. Uncertainty modeling: Medical signals inherently contain uncertainty due to physiological variation, measurement noise, and subjective interpretation. Soft sets naturally accommodate this uncertainty without requiring precise membership functions.
2. Threshold adaptivity: The ATSS and TDSS methods automatically adapt to the statistical properties of each recording, making them robust to inter-subject variability common in medical data.
3. Feature integration: The similarity measures effectively combine information from multiple temporal features, capturing both amplitude and derivative patterns essential for cardiac assessment.
4. Noise resilience: The threshold-based approach is helpful in filtering noise while preserving clinically relevant patterns, as demonstrated by improved specificity compared to traditional methods.

Practical significance. On a subset of 1000 recordings from the PhysioNet Challenge 2016 dataset, our method achieved a challenge score of 0.776, compared to the reported baseline performance of 0.71, representing approximately a 9% increase. This medical application demonstrates that the proposed soft set methodology extends beyond the original climate data domain, suggesting broad applicability to various time series classification problems where uncertainty and noise are prevalent.

- Screening applications: Higher sensitivity (0.780) enables detection of pathological cases in population screening scenarios.

- False positive reduction: Improved specificity (0.773) reduces unnecessary referrals and healthcare costs.
- Resource-limited settings: The computational efficiency of soft set methods makes them suitable for mobile health applications and telemedicine platforms.

The successful application to PhysioNet Challenge data validates the clinical utility of soft set theory and establishes a foundation for broader medical time series analysis applications.

5. Discussion

This study addresses a fundamental gap in the intersection of soft set theory and time series analysis. While extensive research has explored soft set applications in decision-making [17,20], healthcare [22], and network analysis [21], the systematic application to temporal data analysis has remained largely unexplored, despite the natural alignment between soft set characteristics and time series uncertainty modeling. Moreover, unlike the traditional approaches by Majumdar and Samanta [5] and Kharal [6,7] which focus primarily on static soft set comparisons, the introduced ratio-based similarity measure represents a significant advancement over existing soft set similarity measures. The two proposed representation methods, amplitude threshold soft set and time derivative threshold soft set, fill a critical methodological gap. While recent advances in fuzzy time series [31, 33] and neutrosophic temporal modeling [34] have addressed uncertainty in time series, they require membership function definitions or complex neutrosophic operations. The given soft set approach eliminates these requirements. The experimental validation demonstrates several advantages over contemporary approaches. In medical signal processing, while existing methods rely heavily on traditional machine learning with engineered features, the given soft set representation naturally captures qualitative cardiac patterns without requiring domain-specific feature engineering. The parameter-based representation enables long-term pattern analysis without the stationarity assumptions required by conventional statistical approaches [28], while avoiding the membership function subjectivity inherent in fuzzy clustering methods [31]. The proposed framework demonstrates complementary potential with recent advances in uncertainty modeling. The approach could potentially integrate with Pythagorean fuzzy decision-making frameworks [36,37] for enhanced multi-criteria temporal analysis, addressing the multi subject interest coordination challenges identified in recent energy systems evaluation [36].

Several important limitations warrant acknowledgment. The current evaluation on 1000 PhysioNet recordings [39], while statistically significant, represents 28% of the complete corpus. Future validation should encompass the full 3541-recording dataset to strengthen generalizability claims. The binary classification framework, though clinically valuable, could be extended to multi-class schemes distinguishing specific cardiac pathologies, potentially incorporating the advanced uncertainty handling demonstrated in recent multi-attribute decision-making algorithms, such as those studied by [26]. The feature extraction methodology presents enhancement opportunities through domain-specific cardiac parameters integration, including S1/S2 timing relationships and heart rate variability measures. Such extensions could leverage the confidence of soft set approaches [12] for handling varying feature reliability. Computational optimization for real-time applications requires investigation, particularly for continuous cardiac monitoring scenarios. The current implementation demonstrates offline analysis feasibility, but real-time deployment would benefit from parallel processing architectures and algorithmic refinements inspired by recent advances in efficient uncertainty modeling [25, 34, 36, 37].

The successful integration of soft set theory with time series analysis establishes a foundation for broader temporal data applications. The approach shows particular promise for scenarios prioritizing uncertainty modeling and qualitative pattern recognition over precise numerical forecasting, suggesting applications in electroencephalography analysis, respiratory pattern recognition, and financial time series analysis where traditional statistical assumptions may not hold. The applications in both the medical and climatological domains demonstrates the framework's versatility, addressing the application diversity gap identified in recent soft set literature reviews [27]. This cross-domain validation strengthens the argument for soft set theory as a robust uncertainty modeling tool for temporal phenomena.

The soft set extraction methodology presented in this study for time series analysis provides a foundational framework that can be extended to incorporate diverse temporal characteristics. Beyond the amplitude and time-derivative approaches demonstrated, the methodology can be adapted to extract soft sets based on frequency domain features, spectral characteristics, or periodicity patterns inherent in time series data. This flexibility opens new avenues for leveraging the unique capabilities of soft set theory in temporal data interpretation.

We can consider the following future directions for this research: Frequency-based soft set construction: Utilizing spectral features, dominant frequencies, or harmonic content as soft set parameters; temporal pattern recognition: Incorporating seasonality, trend components, or cyclic behaviors into soft set representations; multi-scale analysis: Developing soft sets that capture patterns at different temporal resolutions; domain-specific adaptations: Tailoring soft set extraction methods to specific application domains (biomedical, financial, environmental).

The parameter-based nature of soft sets provides inherent advantages for such extensions, as new temporal characteristics can be naturally incorporated without requiring fundamental methodological changes. This adaptability positions soft set theory as a promising framework for advancing time series analysis in uncertain and complex temporal environments.

6. Conclusions

This study successfully bridges the gap between soft set theory and time series analysis through two key methodological innovations: the amplitude and time-derivative threshold soft set representation methods, and a novel ratio-based similarity measure specifically designed for temporal data analysis. The experimental validation demonstrates the framework's effectiveness across diverse domains with quantifiable results. In medical signal processing, the proposed soft set approach achieved a challenge score of 0.776 ± 0.028 for heart sound classification using the PhysioNet 2016 dataset [39], representing a 9.3 % improvement over the official baseline and a remarkable 20.0% improvement in sensitivity (0.780 vs. 0.65). The ratio-based normalized similarity measure with random forest classification demonstrated superior performance with optimal parameters $\alpha = 0.15$ and $\beta = 0.15$, validating the systematic parameter optimization methodology. Unlike existing fuzzy time series approaches [31, 33] that require subjective membership functions, our parameter-based representation maintains interpretability while effectively handling uncertainty. The systematic comparison of three similarity measures (Jaccard-based, ratio-based raw, and ratio-based normalized) on medical data conclusively demonstrated the superiority of the ratio-based normalized approach.

For climatological analysis, the framework captured 58-year temperature patterns across six Brazilian cities [38], demonstrating long-term temporal pattern recognition capabilities that exceed the scope typically addressed in fuzzy time series literature [32]. The cross-domain validation across both biomedical signal processing and climatological analysis establishes the framework's versatility and broad applicability. The ratio-based similarity measure addresses critical limitations in existing soft set similarity measures [5–7], providing enhanced discriminative capabilities for temporal patterns that static measures cannot adequately capture. This work establishes soft set theory as a viable and competitive alternative for time series analysis where uncertainty modeling and qualitative pattern recognition are prioritized over precise numerical forecasting. The rigorous experimental design, including subject-based stratified splits and 5-fold cross-validation, ensures the reliability and reproducibility of the findings. Future research directions include extension to multivariate time series analysis, integration with emerging uncertainty modeling paradigms such as Pythagorean fuzzy frameworks [36, 37], real-time implementation optimization for continuous cardiac monitoring applications, and expansion to multi-class classification schemes for more granular cardiac pathology detection. The successful validation on 1000 recordings provides a foundation for scaling to the complete 3541-recording PhysioNet corpus.

Use of Generative-AI tools declaration

The author declares he has not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The author declares that he has no conflicts of interest.

References

1. D. Molodtsov, Soft set theory—first results, *Comput. Math. Appl.*, **37** (1999), 19–31. https://doi.org/10.1016/S0898-1221(99)00056-5
2. P. K. Maji, R. Biswas, A. R. Roy, Soft set theory, *Comput. Math. Appl.*, **45** (2003), 555–562. https://doi.org/10.1016/S0898-1221(03)00016-6
3. P. K. Maji, R. Biswas, A. R. Roy, Fuzzy soft sets, *J. Fuzzy Math.*, **9** (2001), 589–602.
4. A. Tversky, Features of similarity, *Psychol. Rev.*, **84** (1977), 327–352. https://doi.org/10.1037/0033-295X.84.4.327
5. P. Majumdar, S. K. Samanta, Similarity measure of soft sets, *New Math. Nat. Comput.*, **4** (2008), 1–12. https://doi.org/10.1142/S1793005708000908
6. N. H. Sulaiman, D. Mohamad, *A Jaccard-based similarity measure for soft sets*, In: 2012 IEEE Symposium on Humanities, Science and Engineering Research, IEEE, 2012, 659–663. https://doi.org/10.1109/SHUSER.2012.6268901
7. A. Kharal, Distance and similarity measures for soft sets, *New Math. Nat. Comput.*, **6** (2010), 321–334. https://doi.org/10.1142/S1793005710001724

8. S. Salsabeela, S. J. John, *A similarity measure of picture fuzzy soft sets and its application*, In: International Conference on Nonlinear Applied Analysis and Optimization, Singapore: Springer, 2023, 381–389. https://doi.org/10.1007/978-981-99-0597-3_26
9. X. Yang, T. Y. Lin, J. Yang, Y. Li, D. Yu, Combination of interval-valued fuzzy set and soft set, *Comput. Math. Appl.*, **58** (2009), 521–527. https://doi.org/10.1016/j.camwa.2009.04.019
10. F. Feng, C. Li, B. Davvaz, M. I. Ali, Soft sets combined with fuzzy sets and rough sets: A tentative approach, *Soft Comput.*, **14** (2010), 899–911. https://doi.org/10.1007/s00500-009-0465-6
11. J. Zhan, Q. Liu, T. Herawan, A novel soft rough set: Soft rough hemirings and corresponding multicriteria group decision making, *Appl. Soft Comput.*, **54** (2017), 393–402. https://doi.org/10.1016/j.asoc.2016.09.012
12. M. Aggarwal, Confidence soft sets and applications in supplier selection, *Comput. Ind. Eng.*, **127** (2019), 614–624. https://doi.org/10.1016/j.cie.2018.11.005
13. S. Vijayabalaji, A. Ramesh, Belief interval-valued soft set, *Expert Syst. Appl.*, **119** (2019), 262–271. https://doi.org/10.1016/j.eswa.2018.10.054
14. N. Hijriati, I. S. Yulianti, D. S. Susanti, D. Anggraini, The construction of soft sets from fuzzy subsets, *Barekeng J. Math. Appl.*, **17** (2023), 1473–1482. https://doi.org/10.30598/barekengvol17iss3pp1473-1482
15. Z. A. Ameen, S. A. Ghour, Cluster soft sets and cluster soft topologies, *Comput. Appl. Math.*, **42** (2023), 337. https://doi.org/10.1007/s40314-023-02476-7
16. L. Fu, F. Qin, Ternary fuzzy soft sets, *Comput. Appl. Math.*, **43** (2024), 253. https://doi.org/10.1007/s40314-024-02757-9
17. N. Çağman, S. Enginoğlu, Soft matrix theory and its decision making, *Comput. Math. Appl.*, **59** (2010), 3308–3314. https://doi.org/10.1016/j.camwa.2010.03.015
18. G. Yaylalı, N. Çakmak Polat, B. Tanay, A soft interval based decision making method and its computer application, *Found. Comput. Decis. S.*, **46** (2021), 273–296. https://doi.org/10.2478/fcds-2021-0018
19. M. A. Balcı, L. M. Batrancea, Ö. Akgüller, Network-induced soft sets and stock market applications, *Mathematics*, **10** (2022), 3964. https://doi.org/10.3390/math10213964
20. G. Yaylalı, N. Çakmak Polat, B. Tanay, A generalized belief interval-valued soft set with applications in decision making, *Soft Comput.*, **26** (2022), 6019–6020. https://doi.org/10.1007/s00500-022-07113-9
21. Ö. Akgüller, A soft set theoretic approach to network complexity and a case study for Turkish Twitter users, *Appl. Soft Comput.*, **143** (2023), 110344. https://doi.org/10.1016/j.asoc.2023.110344
22. D. Gifu, Soft sets extensions: Innovating healthcare claims analysis, *Appl. Sci.*, **14** (2024), 8799. https://doi.org/10.3390/app14198799
23. E. Bayram, G. Çelik, M. Gezek, An advanced encryption system based on soft sets, *AIMS Math.*, **9** (2024), 32232–32256. https://doi.org/10.3934/math.20241547
24. M. Saqlain, P. Kumam, W. Kumam, Neutrosophic linguistic valued hypersoft set with application: Medical diagnosis and treatment, *Neutrosophic Sets Syst.*, **63** (2024), 130–152.

25. M. E. M. Abdalla, A. Uzair, A. Ishtiaq, M. Tahir, M. Kamran, Algebraic structures and practical implications of interval-valued Fermatean neutrosophic super hypersoft sets in healthcare, *Spectrum Oper. Res.*, **2** (2025), 199–218. https://doi.org/10.31181/sor21202523
26. G. Y. Umul, A multi-attribute group decision-making algorithm based on soft intervals that considers the priority rankings of group members on attributes of objects, along with some applications, *AIMS Math.*, **10** (2025), 4709–4746. https://doi.org/10.3934/math.2025228
27. J. C. R. Alcantud, A. Z. Khameneh, G. S. García, M. Akram, A systematic literature review of soft set theory, *Neural Comput. Appl.*, **36** (2024), 8951–8975. https://doi.org/10.1007/s00521-024-09552-x
28. L. Ye, G. Yang, E. V. Ranst, H. Tang, Time-series modeling and prediction of global monthly absolute temperature for environmental decision making, *Adv. Atmos. Sci.*, **30** (2013), 382–396. https://doi.org/10.1007/s00376-012-1252-3
29. A. Aktar, K. C. Roy, Comparative study of changing pattern of temperature for various periods of time, *IOSR J. Math.*, **13** (2017), 6–15. https://doi.org/10.9790/5728-1301050615
30. C. I. Johnpaul, M. V. Prasad, S. Nickolas, G. R. Gangadharan, Fuzzy representational structures for trend based analysis of time series clustering and classification, *Knowl.-Based Syst.*, **222** (2021), 106991. https://doi.org/10.1016/j.knosys.2021.106991
31. M. V. D. S. Ferreira, R. Rios, R. Mello, T. N. Rios, Using fuzzy clustering to address imprecision and uncertainty present in deterministic components of time series, *Appl. Soft Comput.*, **113** (2021), 108011. https://doi.org/10.1016/j.asoc.2021.108011
32. R. M. Pattanayak, H. S. Behera, S. Panigrahi, A novel probabilistic intuitionistic fuzzy set based model for high order fuzzy time series forecasting, *Eng. Appl. Artif. Intel.*, **99** (2021), 104136. https://doi.org/10.1016/j.engappai.2020.104136
33. M. V. D. S. Ferreira, R. Rios, T. N. Rios, sci-FTS: Using soft clustering on intrinsic mode functions to model fuzzy time series, *Softw. Impacts*, **11** (2022), 100230. https://doi.org/10.1016/j.simpa.2022.100230
34. S. A. Edalatpanah, F. S. Hassani, F. Smarandache, A. Sorourkhah, D. Pamucar, B. Cui, A hybrid time series forecasting method based on neutrosophic logic with applications in financial issues, *Eng. Appl. Artif. Intel.*, **129** (2024), 107531. https://doi.org/10.1016/j.engappai.2023.107531
35. J. K. Sana, M. M. Masud, M. S. Rahman, M. S. Rahman, Patient similarity computation for clinical decision support: An efficient use of data transformation, combining static and time series data, *arXiv Preprint*, 2025. https://doi.org/10.48550/arXiv.2506.07092
36. S. Yin, Y. Zhao, A. Hussain, K. Ullah, Comprehensive evaluation of rural regional integrated clean energy systems considering multi-subject interest coordination with Pythagorean fuzzy information, *Eng. Appl. Artif. Intel.*, **138** (2024), 109342. https://doi.org/10.1016/j.engappai.2024.109342
37. R. Li, M. Zhang, S. Yin, N. Zhang, T. Mahmood, Developing a conceptual partner selection framework for matching public–private partnerships of rural energy internet project using an integrated fuzzy AHP approach for rural revitalization in China, *Heliyon*, **10** (2024), e31096. https://doi.org/10.1016/j.heliyon.2024.e31096

38. T. kaggle repository, Temperature time-series for some Brazilian cities, Available: https://www.kaggle.com/datasets/volpatto/temperature-timeseries-for-some-brazilian-cities.
39. PhysioNet, PhysioNet/computing in cardiology challenge 2016, 2016. Available: https://physionet.org/content/challenge-2016/.

AIMS Press

© 2025 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)