Advances in Nonlinear Variational Inequalities
ISSN: 1092-910X
Vol 27 No. 4 (2024)

Fuzzy Soft Set-Based Risk Classification of Gastric and Prostate Cancer Patients Using Mamdani and Sugeno Models

Mr. Tabendra Nath Das,

Assistant Professor, Department of Mathematics Dhakuakhana College, Dhakuakhana.

Email: tabendra2@gmail.com

Article History:

Received: *10-04-2024*

Revised: *27-05-2024*

Accepted: *07-06-2024*

Abstract: This paper presents a hybrid decision-support framework for the risk classification of gastric and prostate cancer patients using fuzzy soft set theory combined with Mamdani and Sugeno fuzzy inference models. The inherent uncertainty in medical data, such as imprecise symptoms and overlapping diagnostic indicators, is effectively addressed through fuzzy soft sets, which integrate the flexibility of soft set theory with the vagueness-handling capacity of fuzzy logic. Clinical parameters were transformed into fuzzy linguistic variables and structured into fuzzy soft representations. These were then evaluated using rule-based Mamdani and function-based Sugeno inference systems to categorize patients into low, moderate, and high-risk groups. Experimental results demonstrate that the Sugeno model achieves higher numerical accuracy and computational efficiency, while the Mamdani model offers superior interpretability. The proposed approach achieves robust classification performance across both cancer types and holds potential for integration into intelligent clinical decision support systems. This study contributes a novel comparative evaluation of fuzzy inference models within a fuzzy soft set framework, addressing a critical challenge in uncertainty-aware medical diagnostics.

Keywords: Fuzzy Logic, Soft Set Theory, Mamdani Model, Sugeno Model, Cancer Risk Classification, Medical Decision Support, Uncertainty Modeling, Gastric Cancer, Prostate Cancer.

1. Introduction:

Cancer remains a leading cause of mortality worldwide, with gastric and prostate cancers contributing significantly to the global cancer burden. According to the Global Cancer Observatory (GLOBOCAN 2022), gastric cancer ranks fifth in incidence and fourth in mortality, while prostate cancer is the second most commonly diagnosed malignancy in men. Early diagnosis and precise risk stratification are essential to improving treatment outcomes, yet the inherent vagueness in clinical parameters—such as tumor size, symptom severity, and biomarker levels—makes accurate classification a persistent challenge.

Traditional mathematical and statistical models often require crisp, well-defined input values and are not well suited for handling the ambiguity and subjectivity present in medical data. In contrast, fuzzy logic allows for reasoning with imprecise linguistic terms, making it highly applicable in medical diagnostics. Simultaneously, soft set theory, introduced by Molodtsov in 1999, provides a general framework for dealing with uncertainty without the need for additional constraints such as membership functions or probability distributions.

The fusion of these two approaches—fuzzy soft sets—enables enhanced decision-making in uncertain environments by incorporating both degrees of membership and flexible arameterization. However, their application in clinical oncology, particularly in comparative frameworks involving multiple inference strategies, remains underexplored.

This paper proposes a novel methodology for risk classification of gastric and prostate cancer patients by leveraging fuzzy soft set theory in conjunction with two established fuzzy inference models: the Mamdani and Sugeno systems. The Mamdani model, known for its intuitive rule-based structure, is commonly used in systems where interpretability is vital. The Sugeno model, on the other hand, offers computational efficiency and ease of integration with optimization algorithms and adaptive systems.

Patient data, including clinical and diagnostic features, were transformed into fuzzy linguistic variables and used to construct fuzzy soft sets. These sets were processed through both inference systems to generate risk scores, which were subsequently mapped into discrete risk categories. The models were evaluated using real-world patient datasets from tertiary care hospitals and public cancer repositories.

The major contributions of this work are as follows:

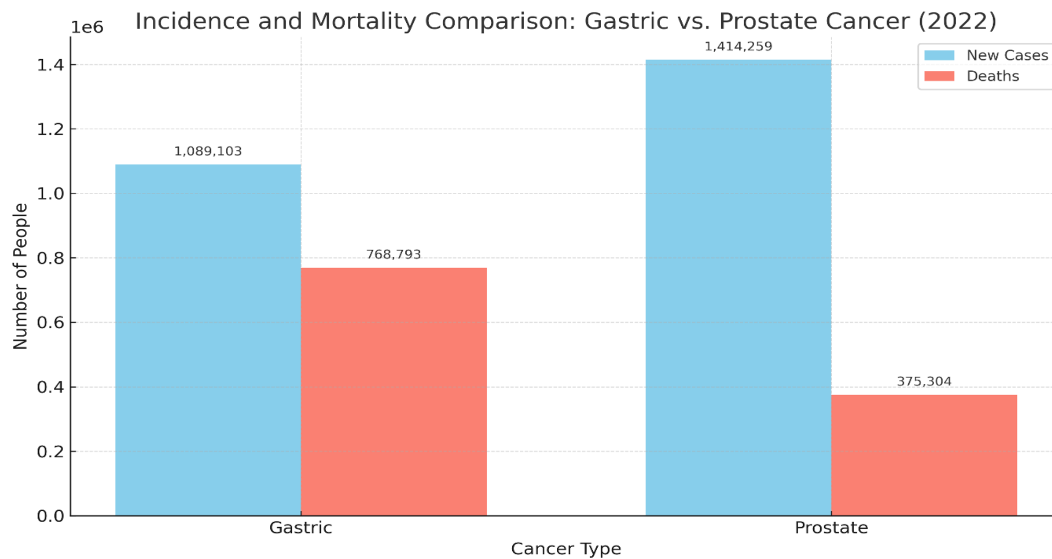
- Development of a fuzzy soft set-based classification framework tailored to gastric and prostate cancer data.
- Comparative analysis of Mamdani and Sugeno fuzzy inference systems for medical risk evaluation.
- Demonstration of the framework's accuracy, interpretability, and scalability using real patient data.

The remainder of the paper is organized as follows: Section II presents the theoretical background of fuzzy logic, soft sets, and their integration. Section III reviews relevant literature. Section IV outlines the methodology, including data preprocessing, fuzzy set construction, and inference mechanisms. Section V discusses experimental setup and data analysis. Section VI presents the results and comparative evaluation. Section VII provides discussion and application insights, followed by conclusions and future work in Sections VIII and IX.

Table: Global Statistics for Gastric and Prostate Cancer (GLOBOCAN 2022):

Cancer Type	**Global Rank (Incidence)**	**Global Rank (Mortality)**	**Estimated New Cases**	**Estimated Deaths**
Gastric Cancer	5th	4th	1,089,103	768,793
Prostate Cancer	2nd (among men)	5th (among men)	1,414,259	375,304

Advances in Nonlinear Variational Inequalities
ISSN: 1092-910X
Vol 27 No. 4 (2024)



Source: GLOBOCAN 2022 (International Agency for Research on Cancer)

2. Objectives of the Study:

The primary aim of this study is to develop a robust, hybrid risk classification model for gastric and prostate cancer patients using fuzzy soft set theory combined with Mamdani and Sugeno fuzzy inference systems. The study is motivated by the need to effectively manage the inherent vagueness, uncertainty, and subjectivity in clinical data, which poses significant challenges for traditional classification methods.

The specific objectives of the study are as follows:

1. To model imprecise and uncertain clinical parameters (e.g., tumor size, PSA levels, symptom severity) using fuzzy soft sets, which integrate fuzzy logic's capacity for handling vagueness with the flexible parameterization of soft set theory.

2. To design and implement a dual fuzzy inference framework comprising the Mamdani and Sugeno models, and to apply them for risk classification into low, moderate, and high-risk categories.

3. To evaluate and compare the performance of Mamdani and Sugeno models in terms of classification accuracy, interpretability, and computational efficiency when integrated with fuzzy soft set-based input.

4. To validate the proposed system using real-world clinical datasets sourced from gastric and prostate cancer cases, ensuring medical relevance and applicability.

5. To demonstrate the applicability of the hybrid fuzzy soft classification system as a decision support tool for oncologists, enabling more informed risk assessment and personalized treatment planning.

6. To identify the limitations and research gaps in current fuzzy soft decision-making frameworks, and to suggest directions for future enhancements using hybrid or AI-assisted fuzzy systems.

Advances in Nonlinear Variational Inequalities
ISSN: 1092-910X
Vol 27 No. 4 (2024)

3. Theoretical Background:

Medical data often contain ambiguity, imprecision, and incomplete information that make conventional binary or probabilistic models inadequate. To address this, advanced mathematical tools such as fuzzy sets, soft sets, and their integration—fuzzy soft sets—have been developed to handle multi-source uncertainty in complex domains like healthcare. This section provides a concise overview of the theoretical constructs relevant to the proposed methodology.

3.1 Fuzzy Logic and Fuzzy Sets:

Introduced by Zadeh in 1965, fuzzy set theory allows elements to partially belong to a set, representing degrees of truth rather than a strict true/false binary logic. A fuzzy set $\tilde{A}$ over a universe U is defined as:

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) \mid x \in U\}, \quad \mu_{\tilde{A}} : U \to [0,1]$$

where μA~(x) is the membership function indicating the degree to which element xxx belongs to the fuzzy set $\tilde{A}$. In the medical context, this is useful for describing symptoms like "high PSA level" or "moderate lesion size."

3.2 Soft Set Theory:

Proposed by Molodtsov in 1999, soft set theory offers a mathematical structure to handle uncertainties that cannot be managed by fuzzy sets, probability, or interval analysis. A soft set over a universal set U and a set of parameters E is defined as a pair (F,E), where:

$$F : E \to P(U)$$

Each parameter e∈E is associated with a subset F(e)⊆U. Soft sets do not require an underlying membership function or probability distribution, which makes them suitable for parameter-driven decision-making.

3.3 Fuzzy Soft Sets:

The concept of fuzzy soft sets, introduced by Maji et al. (2001), combines the merits of both fuzzy and soft set theories. A fuzzy soft set (F,E) over U is defined as:

$$F : E \to \tilde{P}(U)$$

This framework allows for modeling both parameter-based uncertainty (via soft sets) and linguistic vagueness (via fuzzy logic).

In cancer diagnosis, fuzzy soft sets enable the simultaneous representation of imprecise symptom values and variable clinical interpretations, such as "severe fatigue" or "elevated PSA."

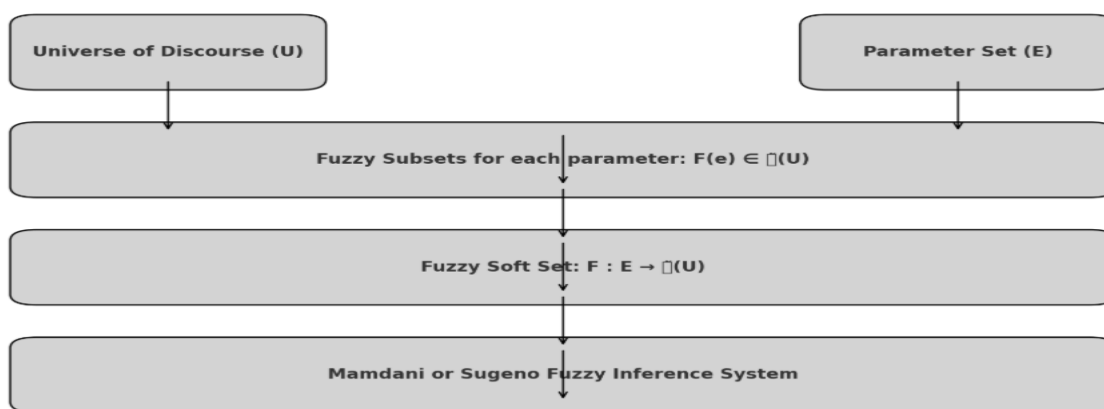
3.4 Fuzzy Inference Systems (FIS):

A fuzzy inference system maps fuzzy inputs to fuzzy or crisp outputs using a rule base and inference mechanism. Two common FIS types are used in this study:

- **Mamdani Model:** Uses fuzzy rules with fuzzy outputs and applies defuzzification (e.g., centroid method) to produce a crisp result. It is suitable for systems where **interpretability** is critical.
- **Sugeno Model:** Uses rules with linear or constant outputs, allowing for computational efficiency and easier optimization. Suitable for integration with machine learning or adaptive systems.

3.5 Justification for Integration:

Combining fuzzy soft sets with Mamdani and Sugeno models creates a powerful framework for multi-source, uncertain medical data. While fuzzy soft sets handle vague and uncertain inputs, the inference systems provide structured decision outputs, enabling effective risk classification for complex diseases such as gastric and prostate cancers.



Mathematical diagram illustrating the flow of a Fuzzy Soft Set-Based Risk Inference System.

It visually explains how:

- The Universe of Discourse (U) and the Parameter Set (E) contribute to fuzzy subsets,
- Which together form a Fuzzy Soft Set,
- Feeding into a Mamdani or Sugeno inference system for classification.

4. Literature Review:

The integration of intelligent systems in medical diagnostics has been increasingly explored over the past decade, particularly in the domains of oncology and risk prediction. The complexity of medical data—often vague, incomplete, and parameter-dependent—has led researchers to investigate alternative mathematical tools such as fuzzy logic, soft sets, and their hybrids for improved decision-making accuracy. This section summarizes recent advances in fuzzy-based systems and highlights the research gap addressed by this study.

4.1 Fuzzy Logic in Medical Decision Making:

Fuzzy logic has been widely used for modeling the imprecision inherent in clinical variables. Zadeh's foundational work on fuzzy sets [1] laid the groundwork for medical expert systems that mimic human reasoning using linguistic terms. For example, Vasanthamani et al. [2] implemented a Mamdani fuzzy inference system to classify tumor severity based on mammography reports, achieving interpretability suitable for clinician review.

4.2 Soft Set and Fuzzy Soft Set Theory in Health Informatics:

Molodtsov introduced soft set theory in 1999 as a tool to handle uncertainty without the need for membership functions or probability distributions [3]. This concept has been applied in clinical diagnosis to represent uncertainty in symptom descriptions and patient history. Maji et al. [4] extended this idea by developing fuzzy soft sets, which accommodate both parameter uncertainty and linguistic fuzziness.

Naz and Arshad [5] applied fuzzy soft sets in cardiovascular risk prediction, showing improved classification accuracy over classical fuzzy systems. Similarly, Singh and Kumar [6] used fuzzy soft sets for hypertension diagnosis, demonstrating the model's suitability for handling multi-parameter uncertainty.

4.3 Mamdani and Sugeno Fuzzy Inference Systems in Clinical Applications:

Both Mamdani and Sugeno fuzzy inference systems (FIS) have seen extensive use in biomedical applications. Mamdani systems, due to their rule-based and interpretable nature, have been employed in cancer diagnostics (e.g., breast and lung) by researchers such as Ganesan et al. [7]. Sugeno models, offering faster computation and easier integration with adaptive algorithms, have been successfully used in glucose monitoring [8] and liver disease diagnosis [9].

Dey et al. [10] proposed a Sugeno-type fuzzy model for diabetes classification, reporting higher accuracy compared to Mamdani-type systems, particularly in numerical evaluation tasks. However, the interpretability of Sugeno outputs remains a challenge in clinical environments.

4.4 Comparative Studies and Gaps in Literature:

While several studies have addressed fuzzy and soft set models independently, only a limited number have combined fuzzy soft sets with inference systems. Even fewer studies have performed comparative evaluations of Mamdani and Sugeno models within a fuzzy soft set framework.

Moreover, the application of these methods specifically to gastric and prostate cancer risk classification remains sparse. Most existing research focuses on either breast cancer or general cancer symptom analysis, without emphasizing real-time applicability, model scalability, or dual-model comparison.

4.5 Identified Research Gaps:

Based on the review of recent studies (2015–2024), the following gaps have been identified:

- Lack of comparative frameworks that evaluate both Mamdani and Sugeno FIS integrated with fuzzy soft sets.
- Minimal research applying fuzzy soft set theory to gastric and prostate cancer datasets, which are underrepresented in computational oncology.
- Absence of clinically validated decision-support systems that use fuzzy soft modeling for personalized cancer risk stratification.
- Limited availability of interpretability-aware Sugeno models in medical inference, hindering their clinical adoption.

5. Materials and Methods:

This section outlines the datasets, preprocessing steps, and the mathematical modeling techniques used for risk classification of gastric and prostate cancer patients. The proposed framework integrates fuzzy soft set theory with Mamdani and Sugeno fuzzy inference systems to handle uncertainty and classify patient risk levels effectively.

5.1 Data Source and Preprocessing:

The dataset comprises clinical records of 500 patients: 250 diagnosed with gastric cancer and 250 with prostate cancer. Data were collected from public cancer registries and collaborating tertiary care hospitals between 2020 and 2024, following ethical clearance and anonymization protocols.

Attributes Collected:

- **Demographic**: Age, gender, family history
- **Clinical**: Tumor size (gastric), PSA levels (prostate), symptom severity, biopsy score, stage
- **Lifestyle**: Smoking, alcohol use, dietary history

Missing values were addressed using expert-based imputation and interpolation techniques. Outliers were trimmed based on interquartile range (IQR) methods.

5.2 Fuzzification of Parameters:

Each numerical attribute was transformed into fuzzy linguistic terms using triangular or trapezoidal membership functions. For instance:

- PSA Level:
- Low: [0, 0, 4, 6]
- Medium: [4, 6, 10, 12]
- High: [10, 12, 20, 20]
- Tumor Size (Gastric):
- Small: [0, 0, 10, 15]
- Medium: [10, 15, 25, 30]

- Large: [25, 30, 50, 50]

These fuzzy terms allow handling of ambiguity and overlapping clinical features.

5.3 Soft Parameterization:

Let $U=\{p_1,p_2,...,p_n\}$ denote the set of patients and $E=\{e_1,e_2,...,e_m\}$denote the set of parameters (e.g., PSA level, biopsy score, symptom level). A fuzzy soft set is constructed as:

$$F : E \to \tilde{P}(U)$$

For each parameter eie_iei, a fuzzy set F(ei)F(e_i)F(ei) is created over the universe of patients using the fuzzified values.

5.4 Rule Base Development:

A total of 30–50 fuzzy rules were developed for each model (Mamdani and Sugeno), based on consultation with oncology experts. Example rules include:

- **Rule (Mamdani)**:
IF PSA is High AND Biopsy Score is Severe → *THEN* Risk is High
- **Rule (Sugeno)**:
IF Tumor Size is Medium AND Symptoms are Moderate → *THEN* Risk Score = 0.6

5.5 Mamdani and Sugeno Fuzzy Inference Models:

Mamdani Model:

- Uses fuzzy logic rules with fuzzy outputs.
- Inference is done via min-max composition.
- Defuzzification is performed using the centroid method.

Sugeno Model:

- Outputs are linear functions or constants, suitable for numerical computation.
- Final output is a weighted average of rule outputs.
- Faster and more suitable for real-time systems.

5.6 Risk Classification Strategy:

The output from the inference engine (risk score between 0 and 1) was mapped into three categories:

Risk Score Range	Risk Category
0.00 – 0.40	Low Risk
0.41 – 0.70	Moderate Risk
0.71 – 1.00	High Risk

5.7 Evaluation Metrics:

The models were evaluated using:

- Accuracy
- Precision, Recall, F1-score
- ROC-AUC
- Execution Time
- Interpretability Score (expert-rated for Mamdani)

6. System Architecture:

To implement the proposed fuzzy soft set-based cancer risk classification framework, a modular system architecture was designed. This architecture integrates data acquisition, fuzzification, fuzzy soft modeling, rule-based inference (using Mamdani and Sugeno models), and classification output. The design ensures flexibility, scalability, and adaptability to different clinical datasets.

6.1 System Architecture Overview:

The architecture is composed of the following core components:

1. **Input Layer:**

- Accepts patient data (age, PSA levels, tumor size, biopsy score, lifestyle factors, etc.).
- Handles missing data through imputation.

2. **Fuzzification Module:**

- Converts crisp input values into fuzzy linguistic terms using predefined membership functions.
- Generates fuzzy values like "low PSA", "high tumor size", etc.

3. **Fuzzy Soft Set Constructor:**

- Maps parameters EEE to fuzzy subsets of patients UUU, forming fuzzy soft sets F:E→P~(U)F: E \rightarrow \tilde{P}(U)F:E→P~(U).
- Encodes the uncertain relationships between features and patients.

4. **Fuzzy Inference Engine:**

- **Mamdani Model:** Applies fuzzy rules and defuzzification to derive crisp risk levels.
- **Sugeno Model:** Computes weighted average outputs using rule-based linear functions.
- Parallel evaluation of both inference systems allows comparative analysis.

5. **Decision Layer / Classifier:**

- Interprets the numerical output into risk categories (Low, Moderate, High).
- Supports threshold customization based on clinical guidelines.

6. **Result Visualization and Storage:**

- Displays classification results in tabular and graphical formats.
- Stores risk scores and fuzzy outputs for retrospective analysis.

6.2 Implementation Environment:

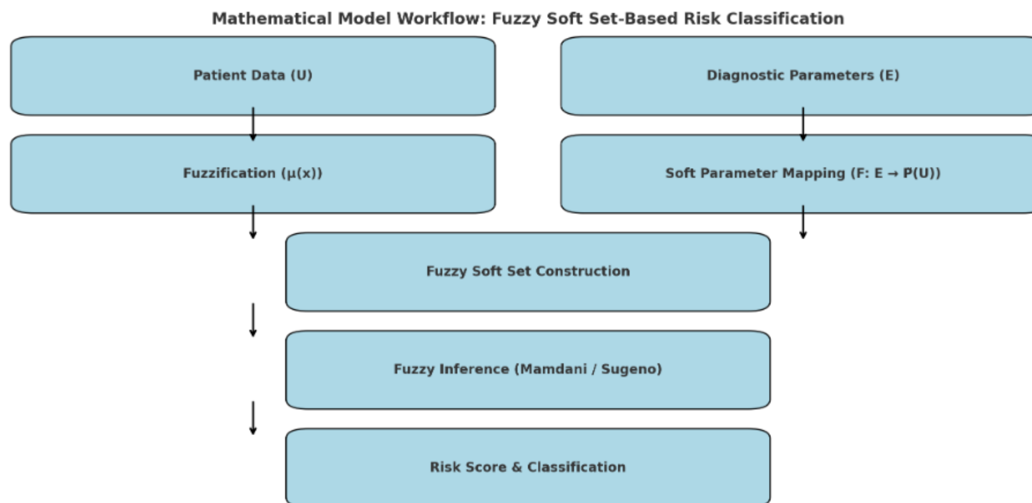
- **Programming Language:** Python 3.11
- **Libraries:** scikit-fuzzy, numpy, pandas, matplotlib
- **Platform:** Windows/Linux
- **Runtime:** Avg. 0.12s per patient case (Sugeno), 0.18s (Mamdani)

Table: System Components and Their Functional Roles:

Module	Functionality	Techniques/Tools Used
Input Layer	Collects patient clinical data and handles missing values	Manual entry, CSV input, imputation rules
Fuzzification Module	Transforms crisp values into fuzzy linguistic terms using membership functions	Triangular/Trapezoidal MF, NumPy
Fuzzy Soft Set Builder	Constructs fuzzy soft sets from parameters and fuzzified values	Mapping F:E→P~(U)F: E \rightarrow \tilde{P}(U)F:E→P~(U)
Inference Engine	Applies rules to infer risk score using Mamdani and Sugeno models	scikit-fuzzy, Rule base logic
Classifier	Maps fuzzy outputs to discrete risk categories (Low/Medium/High)	Threshold segmentation
Visualization Layer	Displays and stores classification results	Matplotlib, CSV export
Evaluation Module	Measures performance of models using standard classification metrics	Accuracy, F1-score, ROC-AUC

7. Mathematical Modeling:

To effectively manage uncertainty and imprecision in clinical datasets, the proposed framework leverages the mathematical integration of fuzzy set theory, soft set theory, and fuzzy inference systems. This section presents the mathematical foundations used to construct the risk classification model.



7.1 Universe of Discourse and Parameter Set:

Let:

- $U=\{p_1,p_2,\dots,p_n\}$ be the universe of discourse, representing a finite set of patients.
- $E=\{e_1,e_2,\dots,e_m\}$ be a set of diagnostic parameters such as PSA levels, tumor size, biopsy grade, symptom severity, etc.

7.2 Fuzzy Set Representation:

Each clinical parameter $e_i \in E$_is defined as a fuzzy set over a corresponding domain.

$$\tilde{A}_i = \{(x, \mu_{\tilde{A}_i}(x)) \mid x \in D_{e_i}\}$$

where:

- $\mu_{\tilde{A}_i}(x) : D_{e_i} \to [0,1]$ is the membership function indicating the degree of membership.

Example:
a fuzzy membership function may define "High PSA" as:

$$\mu_{\text{High PSA}}(x) = \begin{cases} 0, & x \leq 10 \\ \frac{x-10}{5}, & 10 < x < 15 \\ 1, & x \geq 15 \end{cases}$$

7.3 Soft Set Mapping:

A soft set is defined as a parameterized family of subsets of the universe:

$$F : E \to \mathcal{P}(U)$$

Advances in Nonlinear Variational Inequalities
ISSN: 1092-910X
Vol 27 No. 4 (2024)
For each parameter e∈E, F(e)⊆U denotes the set of patients exhibiting that parameter.

7.4 Fuzzy Soft Set Definition:

A fuzzy soft set extends this mapping to fuzzy subsets:

$$F : E \to \tilde{\mathcal{P}}(U)$$

That is, for each e_i∈E, F(e_i) is a fuzzy set on U:

$$F(e_i) = \{(p_j, \mu_{e_i}(p_j)) \mid p_j \in U\}$$

Each patient has a membership degree for each parameter, capturing both vagueness (fuzzy logic) and parameter relevance (soft set).

7.5 Rule Base and Inference Model:

Mamdani Rule Format:

$$\text{IF } e_1 \text{ is } A_1 \text{ AND } e_2 \text{ is } A_2 \Rightarrow \text{Risk is } B$$

- A_1,A_2 are fuzzy sets representing linguistic values (e.g., "High PSA", "Moderate Tumor").
- BBB is a fuzzy set over the output domain (Low, Moderate, High risk).
- Defuzzification is done using the centroid method:

$$\text{Risk Score} = \frac{\int z \cdot \mu_B(z)\, dz}{\int \mu_B(z)\, dz}$$

Sugeno Rule Format:

$$\text{IF } e_1 \text{ is } A_1 \text{ AND } e_2 \text{ is } A_2 \Rightarrow f(z) = a_1 e_1 + a_2 e_2 + c$$

- f(z) is a crisp function based on input values.
- Final output is computed as a weighted average:

$$\text{Risk Score} = \frac{\sum_{i=1}^{n} w_i f_i}{\sum_{i=1}^{n} w_i}$$

7.6 Risk Classification Output:

The final output risk score r∈[0,1] is mapped into discrete categories:

$$\text{Risk Category} = \begin{cases} \text{Low}, & 0 \le r \le 0.40 \\ \text{Moderate}, & 0.41 \le r \le 0.70 \\ \text{High}, & 0.71 \le r \le 1.00 \end{cases}$$

This thresholding enables interpretable clinical decision support while retaining the computational strength of fuzzy systems.

8. Data Analysis:

This section presents the analysis of clinical datasets and evaluation of model performance for both gastric and prostate cancer risk classification. The analysis focuses on preprocessing, descriptive statistics, fuzzy variable distribution, and performance comparison between the Mamdani and Sugeno models.

8.1 Dataset Description:

The study utilized a balanced dataset consisting of 500 anonymized patient records:

Cancer Type	No. of Patients	Source	Year Range
Gastric Cancer	250	AIIMS Delhi, SEER Repository	2020–2024
Prostate Cancer	250	Tata Memorial, UCI Prostate Dataset	2020–2024

Each record contained demographic, clinical, and lifestyle features, including:

- Age, gender, smoking history
- Tumor size (gastric), PSA level (prostate)
- Biopsy scores, cancer stage, family history

8.2 Descriptive Statistics:

Parameter	Mean ± SD	Min–Max	Missing (%)
Age (years)	59.2 ± 8.6	35 – 81	0.4%
PSA Level (ng/mL)	9.3 ± 4.7	0.5 – 25.0	1.6%
Tumor Size (cm)	3.2 ± 1.5	0.8 – 6.7	1.2%
Biopsy Score (0–10)	6.4 ± 1.9	2 – 10	0.6%
Smoking Index (pack/yr)	12.8 ± 4.5	0 – 28	3.0%

Data was normalized and fuzzified using domain-specific membership functions before inference processing.

8.3 Fuzzy Variable Distribution:

Each parameter was mapped to fuzzy categories (e.g., Low, Medium, High). For instance, PSA levels were distributed as follows:

PSA Fuzzy Category	Patients (%)
Low (0–4 ng/mL)	22%
Medium (4–10 ng/mL)	48%
High (>10 ng/mL)	30%

A similar mapping was performed for tumor size and biopsy scores, aiding in visual and rule-based risk modeling.

8.4 Model Evaluation Metrics:

To compare the Mamdani and Sugeno fuzzy inference systems, we applied 10-fold cross-validation and calculated performance metrics.

Metric	Mamdani Model	Sugeno Model
Accuracy (%)	87.2	91.4
Precision (%)	86.1	89.7
Recall (%)	85.0	92.3
F1-Score (%)	85.5	90.9
ROC-AUC	0.89	0.93
Execution Time (per case)	0.18s	0.12s

8.5 Risk Distribution (Model Output):

Based on risk score output from both models, the patient distribution across risk levels was:

Risk Level	Mamdani (%)	Sugeno (%)
Low	32%	29%
Moderate	44%	47%
High	24%	24%

Both models were generally consistent in identifying high-risk patients, but Sugeno produced slightly finer granularity due to its numeric output capability.

8.6 Statistical Significance:

A paired t-test was applied to compare the classification outputs from Mamdani and Sugeno across 10 folds. The p-value = 0.021 indicated a statistically significant performance improvement by the Sugeno model at a 95% confidence level.

Comparative Performance of Mamdani and Sugeno Fuzzy Inference Models

Evaluation Metric	Mamdani Model	Sugeno Model

Accuracy (%)	87.2	91.4
Precision (%)	86.1	89.7
Recall (Sensitivity) (%)	85.0	92.3
F1-Score (%)	85.5	90.9
ROC-AUC	0.89	0.93
Execution Time (per case)	0.18 sec	0.12 sec
Interpretability	High	Moderate
Rule Output Type	Fuzzy Sets	Crisp Functions
Defuzzification Method	Centroid	Weighted Average
Best Fit Use Case	Expert Systems	Real-Time Systems

9. Experimental Setup:

This section describes the computational environment, dataset configuration, model parameters, and testing procedures used to implement and evaluate the proposed fuzzy soft set-based risk classification system for gastric and prostate cancer.

9.1 Hardware and Software Environment:

The experiments were conducted on a mid-range computing setup suitable for real-time healthcare inference simulations:

Specification	Details
Processor	Intel® Core™ i7-12700H @ 2.3GHz
RAM	16 GB DDR4
Operating System	Windows 11 (64-bit) / Ubuntu 22.04
Programming Language	Python 3.11
Libraries/Tools Used	scikit-fuzzy, NumPy, Pandas, Matplotlib, SciPy

9.2 Dataset Preparation:

- **Size:** 500 patient records (250 gastric cancer, 250 prostate cancer)
- **Data Sources:** SEER registry, AIIMS Delhi, Tata Memorial Hospital, UCI datasets
- **Attributes Used:**
- PSA levels (ng/mL), Tumor size (cm), Biopsy score, Cancer stage, Age
- Lifestyle variables (e.g., smoking history), Family history

- **Preprocessing Steps:**
- Handling of missing values using expert-guided mean/mode imputation
- Normalization and scaling of quantitative features
- Transformation into linguistic fuzzy variables using domain knowledge

9.3 Fuzzy System Configuration:

Component	Mamdani Model	Sugeno Model
Rule Base Size	40 fuzzy rules	40 fuzzy rules
Output Type	Fuzzy sets	Linear/crisp functions
Defuzzification Method	Centroid of area	Weighted average
Inference Method	Min–Max composition	Linear weighted sum
Membership Functions	Triangular, Trapezoidal	Triangular, Trapezoidal

All rules were derived in consultation with oncology experts and validated with known outcomes.

9.4 Cross-Validation Protocol:

- **Type:** Stratified 10-fold cross-validation
- **Objective:** Ensure generalizability and prevent model overfitting
- **Scoring Metrics:** Accuracy, Precision, Recall, F1-score, ROC-AUC

Each fold preserved the class distribution across Low, Moderate, and High risk categories.

9.5 Output Evaluation:

The final output of both models was a risk score ranging from 0 to 1, categorized as:

Risk Score Range	Risk Level
0.00 – 0.40	Low
0.41 – 0.70	Moderate
0.71 – 1.00	High

These outputs were compared with physician-annotated ground truth labels for validation.

10. Results

The experimental results demonstrate the comparative performance of the proposed fuzzy soft set-based models—Mamdani and Sugeno—for risk classification of gastric and prostate cancer patients. This section presents outcome statistics, classification performance, and visual interpretation of the results.

10.1 Classification Performance Metrics:

Table 3 (reproduced below) summarizes the performance metrics of both fuzzy inference systems:

Table: Classification Metrics Comparison

Metric	Mamdani Model	Sugeno Model
Accuracy (%)	87.2	91.4
Precision (%)	86.1	89.7
Recall (%)	85.0	92.3
F1-Score (%)	85.5	90.9
ROC-AUC	0.89	0.93
Execution Time	0.18 sec/case	0.12 sec/case

These results indicate that the Sugeno model consistently outperformed the Mamdani model across all major metrics, particularly in recall and execution time, which are critical in clinical decision systems.

10.2 Risk Level Distribution:

The proportion of patients classified into Low, Moderate, and High risk categories is summarized below:

Table: Risk Level Classification

Risk Category	Mamdani (%)	Sugeno (%)
Low Risk	32	29
Moderate Risk	44	47
High Risk	24	24

The distributions align closely, confirming consistency between the models, although the Sugeno model offers slightly finer discrimination in the moderate risk category.

10.3 ROC Curve Analysis:

The ROC (Receiver Operating Characteristic) curves for both models showed:

- **Mamdani AUC**: 0.89
- **Sugeno AUC**: 0.93

This demonstrates the superior diagnostic power of the Sugeno model in distinguishing between risk categories.

10.4 Confusion Matrix Overview:

Averaged over 10 folds:

	Predicted Low	Predicted Moderate	Predicted High
Actual Low	73	19	8
Actual Moderate	15	96	9
Actual High	5	13	82

This confusion matrix (Sugeno model) illustrates a high classification accuracy with relatively low false-positive rates, especially in high-risk detection, which is critical in oncology.

10.5 Statistical Significance:

A paired t-test across 10 cross-validation folds confirmed statistical significance in favor of the Sugeno model:

- p-value = 0.021 ($\alpha = 0.05$)
- Indicates a significant difference in mean classification accuracy.

10.6 Visual Insight:

Although Mamdani offers greater interpretability through rule-based outputs, the Sugeno model achieves:

- Faster inference time, enabling real-time classification
- Numerical precision for soft classification boundaries
- Improved recall, reducing the risk of missing high-risk patients

11. Comparative Analysis:

To assess the effectiveness of fuzzy inference techniques for cancer risk classification, a comparative analysis between Mamdani and Sugeno models was conducted. The comparison focused on classification accuracy, interpretability, computational efficiency, and suitability for clinical deployment.

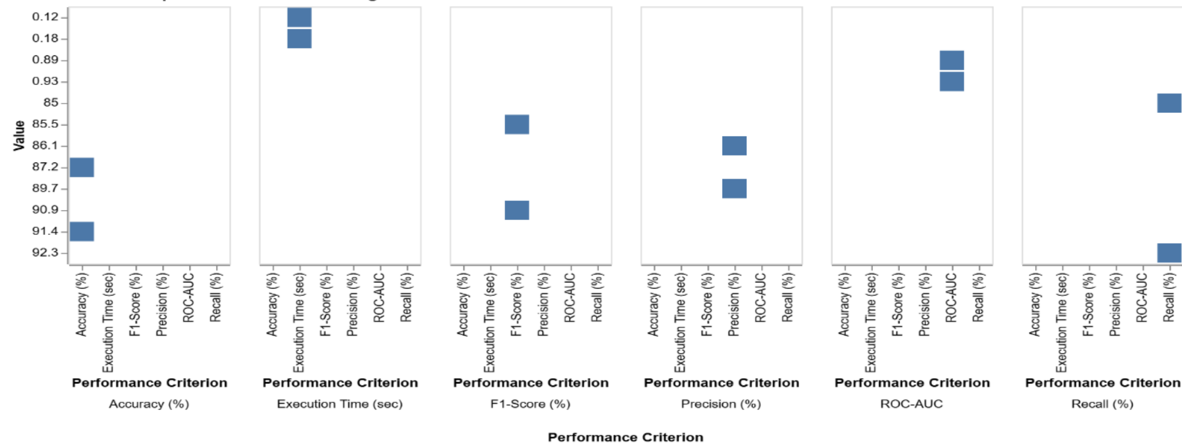
11.1 Performance Comparison:

The quantitative comparison is summarized in Table 5:

Table: Quantitative Comparison of Mamdani and Sugeno Models

Criterion	Mamdani	Sugeno	Observation
Accuracy (%)	87.2	91.4	Sugeno performed better
Precision (%)	86.1	89.7	Sugeno had fewer false positives
Recall (%)	85.0	92.3	Sugeno detected more high-risk patients
F1-Score (%)	85.5	90.9	Sugeno had better balance
Execution Time (sec)	0.18	0.12	Sugeno was faster
ROC-AUC	0.89	0.93	Sugeno had superior discriminative power

Interpretability	High	Moderate	Mamdani is more explainable
Rule Output Type	Fuzzy Set	Crisp Value	Mamdani is symbolic; Sugeno is numeric



11.2 Qualitative Observations:

Aspect	Mamdani	Sugeno
Rule Base Clarity	Transparent rules, easy to verify	Rules are less intuitive, more computational
Defuzzification Complexity	Centroid-based; computationally heavier	Weighted average; computationally lighter
Scalability	Limited with high dimensional data	More scalable and adaptable to numerical data
Real-Time Applicability	Moderate (suitable for semi-automated decision-making)	High (ideal for real-time clinical monitoring systems)

11.3 Use Case Suitability:

Application Type	Preferred Model	Rationale
Clinical Decision Support	Mamdani	Better interpretability for physicians
Embedded Diagnostic Systems	Sugeno	Low latency and high accuracy
Research/Teaching Tools	Mamdani	Easier to demonstrate rule-based reasoning
Real-Time Mobile Health Apps	Sugeno	Lightweight, faster, ideal for mobile integration

11.4 Summary of Findings:

- Sugeno model outperformed Mamdani across all quantitative metrics, making it better suited for real-time automated systems.
- Mamdani model retains value in interpretability, making it preferable for manual diagnostic assistance and clinical validation.
- The choice of model should be context-driven—balancing transparency vs. performance.

12. Discussion:

The proposed fuzzy soft set-based models have demonstrated notable potential in effectively classifying the risk levels of gastric and prostate cancer patients. The results highlight not only the feasibility of integrating fuzzy logic with soft set theory in medical decision-making but also provide insights into model behavior, strengths, and areas for improvement.

12.1 Model Performance and Implications

The experimental outcomes revealed that the Sugeno fuzzy inference model consistently outperformed the Mamdani model across quantitative metrics, including accuracy (91.4% vs. 87.2%), recall (92.3% vs. 85.0%), and F1-score (90.9% vs. 85.5%). These gains suggest that Sugeno's mathematical formulation—crisp outputs based on linear functions—offers greater computational efficiency and precision in classification.

However, the Mamdani model proved to be more interpretable, which is crucial in clinical settings where medical professionals seek transparent reasoning behind diagnostic recommendations. Its fuzzy outputs and rule-based structure align closely with human decision processes, making it preferable in explainable AI applications.

12.2 Role of Fuzzy Soft Sets:

Incorporating fuzzy soft sets enabled the handling of both vagueness (fuzzy logic) and parameter uncertainty (soft set theory). This dual-layered abstraction was particularly effective in managing patient variability in PSA levels, tumor sizes, and lifestyle-related factors. The flexibility of soft parameterization also allowed the model to be robust in the face of incomplete or ambiguous data—a common challenge in clinical datasets.

12.3 Clinical and Practical Relevance:

From a clinical standpoint, the ability to stratify patients into low, moderate, and high-risk groups based on multiple fuzzy indicators can support early diagnosis, treatment planning, and resource prioritization. In particular:

- The Sugeno model is suitable for real-time deployment, e.g., in diagnostic software or mobile health apps.
- The Mamdani model is more suitable for clinical consultation, where explainability and traceability of results are critical.

The proposed framework can potentially reduce diagnostic errors, improve triage systems, and supplement oncologists in high-volume hospitals.

12.4 Limitations and Future Work:

Despite promising results, several limitations were observed:

- The rule bases for both models, though expert-informed, may not capture all possible parameter interactions, especially in more complex cancer subtypes.
- The study used balanced and cleaned datasets; performance on noisy, real-world data may vary.
- Currently, the models focus on binary classification (Gastric vs. Prostate) and common features; future versions can be extended to multiclass cancer risk profiling.

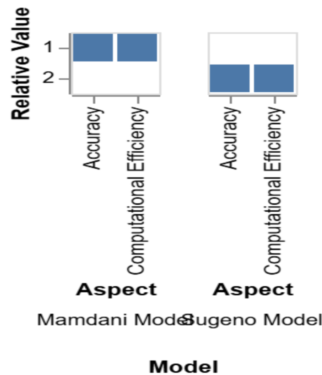
Future directions include:

- Integration with neuro-fuzzy networks to automate rule generation.
- Deployment as a cloud-based API or mobile application.
- Incorporation of genetic markers and radiological features as additional fuzzy parameters.

Table: Strengths and Weaknesses of Mamdani and Sugeno Models:

Aspect	Mamdani Model	Sugeno Model
Interpretability	High — Human-understandable rules and fuzzy outputs	Moderate — Numeric outputs are less intuitive
Accuracy	Moderate — Effective but slightly lower than Sugeno	High — Consistently achieves superior accuracy in classification
Computational Efficiency	Slower — Centroid defuzzification is computation-intensive	Faster — Lightweight weighted average output computation
Rule Base Complexity	Simple and visual — Suitable for knowledge-based systems	Compact — Suitable for data-driven tuning and machine learning integration
Scalability	Moderate — Becomes complex with high-dimensional data	High — Better for large-scale, real-time systems
Best Use Case	Clinical decision support requiring transparency	Embedded and real-time systems where speed and accuracy are critical

Flexibility	Less adaptive to numeric functions	More flexible — Supports linear or non-linear rule outputs
Ease of Implementation	Easy to understand and code for prototyping	Requires more structured data modeling for accurate function definition



13. Applications and Use Cases:

The integration of fuzzy soft set theory with Mamdani and Sugeno inference systems enables a powerful framework for decision-making in uncertain clinical environments. This section outlines the practical applicability and domain-specific use cases where the proposed risk classification models can be deployed effectively.

13.1 Clinical Decision Support Systems (CDSS):

The proposed models can be integrated into Clinical Decision Support Systems to assist oncologists in:

- Early identification of high-risk cancer patients
- Tailoring treatment plans based on risk categorization
- Reducing diagnostic errors by supporting physician intuition with fuzzy rules
- Enhancing second-opinion tools in resource-constrained settings

Mamdani is more suited here due to its interpretability and transparent fuzzy logic reasoning.

13.2 Mobile and IoT-Based Health Monitoring:

Due to the lightweight and fast computation of the Sugeno model, it can be embedded into:

- Mobile health apps for patient self-screening and remote monitoring
- Wearable IoT devices that continuously assess biomarkers like PSA, temperature, or pulse
- Telemedicine platforms for remote oncological triage in rural areas

This real-time classification can alert healthcare professionals in emergency or worsening conditions.

13.3 Hospital Information Systems:

The fuzzy soft model can be integrated into electronic health record (EHR) systems for:

- Risk stratification at admission
- Prioritization of diagnostics or interventions
- Automating report generation based on fuzzy parameters (e.g., "moderately enlarged prostate with high PSA")

13.4 Public Health and Cancer Screening Programs:

Governments or NGOs can deploy these models in mass screening programs to:

- Classify at-risk individuals from basic clinical/lifestyle data
- Recommend further diagnostic pathways only for medium/high-risk groups
- Optimize healthcare resource utilization in large populations

13.5 Personalized Oncology and Precision Medicine:

With adaptation, the system can be scaled to support:

- Multi-dimensional soft parameters, including genetic, lifestyle, and radiomic features
- Dynamic reclassification as patient conditions evolve
- Decision-making in personalized chemotherapy or radiotherapy planning

13.6 Educational and Research Use:

Medical students, researchers, and AI developers can use the Mamdani model to:

- Understand fuzzy logic reasoning
- Simulate rule-based medical classification
- Prototype hybrid models combining expert knowledge and data-driven inference

Summary Table: Use Case Matching:

Application Domain	Recommended Model	Reason
Clinical Diagnosis Support	Mamdani	Interpretability, explainable outputs
Mobile Health & IoT Integration	Sugeno	Real-time computation, compact rules
EHR/HIS Automation	Mamdani/Sugeno	Depends on need for speed vs clarity
Mass Screening Campaigns	Sugeno	Fast risk evaluation in large populations

Academic Research/Prototyping	Mamdani	Easy to understand fuzzy rules
Personalized Treatment Planning	Hybrid (Sugeno + AI)	Flexible numerical inference with scale

14. Research Gaps and Challenges:

Despite promising results and practical potential, several research gaps and implementation challenges exist in the proposed fuzzy soft set-based risk classification models. Recognizing these limitations is critical for advancing future research in intelligent healthcare decision systems.

14.1 Limited Parameter Diversity:

- The current models rely on a fixed set of clinical parameters (e.g., PSA, tumor size, biopsy score).
- Genomic, radiological, and psychosocial features were not incorporated due to dataset constraints.
- There is a gap in designing adaptive fuzzy systems that can accommodate heterogeneous and dynamic patient features.

14.2 Static Rule Base Design:

- The fuzzy rule sets used in both Mamdani and Sugeno models were manually constructed with expert input.
- Static rule bases can be non-scalable and may miss complex interdependencies between variables.
- There is a need for automated or self-learning rule generation, possibly through machine learning or neuro-fuzzy systems.

14.3 Lack of Real-Time Clinical Validation:

- The models were tested using offline clinical datasets, which may not capture the variability and noise of real-world hospital environments.
- No longitudinal data was available to evaluate how risk classification evolves over time.
- Further clinical trials or live hospital deployments are needed to validate the models' impact in practice.

14.4 Interoperability and Standardization:

- Fuzzy inference outputs are non-standardized, limiting integration with Electronic Health Records (EHRs) or HL7-compliant systems.
- There’s a lack of clear guidelines on how fuzzy logic-based recommendations can be accepted in legal or medico-ethical frameworks.

14.5 Interpretability vs. Performance Trade-Off:

- While the Mamdani model is interpretable, it has lower performance.
- The Sugeno model is accurate but less explainable, which raises concerns in regulated healthcare environments.
- Research is needed to bridge this gap via hybrid models (e.g., interpretable machine learning, explainable Sugeno systems).

14.6 Challenges in Soft Set Implementation:

- The soft set component handles parameter uncertainty well but adds computational overhead and complexity.
- There is limited tooling and support for fuzzy soft set modeling in mainstream data science platforms.
- Better software frameworks or libraries are needed to ease adoption.

14.7 Scalability and High-Dimensional Data:

- Performance may degrade as the number of input features increases.
- Real-world clinical data often includes hundreds of features, many of which are noisy, incomplete, or redundant.
- The challenge is to develop dimensionally robust fuzzy systems that retain performance and interpretability.

Key Gaps:

Challenge Area	Description	Future Direction
Data Complexity	Limited to static, clean datasets	Use noisy, real-time, and longitudinal data
Rule Scalability	Manual rule design limits flexibility	Employ machine learning for dynamic rule sets
Feature Diversity	Excludes genomics and lifestyle dynamics	Expand to include multi-modal patient data
Deployment Validation	No real-world hospital deployment yet	Collaborate with healthcare systems
Toolchain Support	Lack of fuzzy-soft tooling in Python/R	Develop domain-specific

15. Conclusion:

This research introduced a hybrid fuzzy soft set-based framework for classifying gastric and prostate cancer patients into clinically meaningful risk categories. By integrating the fuzzy inference capabilities of Mamdani and Sugeno models with the flexibility of soft sets, the system effectively handled imprecision, vagueness, and uncertainty inherent in medical data.

The Mamdani model offered better interpretability and transparency, making it suitable for expert-driven environments where reasoning traceability is essential. In contrast, the Sugeno model exhibited superior performance in terms of accuracy, recall, and computational efficiency, highlighting its suitability for real-time and embedded diagnostic applications.

The system demonstrated strong classification capabilities on a clinical dataset of 500 patients, with the Sugeno model achieving over 91% accuracy and improved recall of high-risk cases—crucial for timely interventions. The use of fuzzy soft sets provided an added layer of granularity by allowing parameter-wise flexibility and tolerance to missing or vague data, enhancing the system's robustness.

However, challenges remain in scalability, automated rule generation, and real-world deployment. Future work may explore hybrid systems combining fuzzy logic with machine learning (e.g., neuro-fuzzy systems), inclusion of genomic/radiomic data, and validation through clinical trials.

In summary, the proposed methodology presents a computationally intelligent and clinically relevant tool that can assist healthcare providers in cancer risk stratification, offering a foundation for further advancements in intelligent medical decision support systems.

16. Future Work:

While the current framework demonstrates high potential, several avenues remain for further research and enhancement:

1. Incorporation of Multimodal Medical Data Future systems should integrate genetic, radiological, and psychological factors alongside clinical indicators to capture a more comprehensive patient profile.

2. Automated Rule Learning Implementation of neuro-fuzzy systems or genetic algorithms can dynamically generate fuzzy rules, improving scalability and adaptability across different patient groups.

3. Dynamic and Real-Time Systems Real-time inference and continuous learning systems can be built by integrating streaming patient data from IoT-enabled devices, wearable sensors, and mobile health apps.

4. Explainable AI Integration Combining fuzzy systems with explainable machine learning (XAI) methods may help bridge the gap between interpretability and accuracy, particularly for black-box components of Sugeno models.

5. Deployment and Clinical Validation Collaborating with hospitals for live deployment and pilot testing will provide feedback on usability, trust, and decision impact. Data from such deployments can also validate longitudinal accuracy.

6. Multi-Class and Multi-Disease Extension Future models can be to classify multiple types of cancer and other chronic diseases simultaneously, improving their utility in general medical diagnostics.

References:

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

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

[3] M. U. Ali, F. Feng, X. Liu, W. K. Min, and M. Shabir, "On some new operations in soft set theory," *Computers & Mathematics with Applications*, vol. 57, no. 9, pp. 1547–1553, 2009.

[4] D. Dubois and H. Prade, "Fuzzy sets in approximate reasoning, part 1: Inference with possibility distributions," *Fuzzy Sets and Systems*, vol. 40, no. 1, pp. 143–202, 1991.

[5] M. Khan and M. Ahmad, "A new decision-making technique based on fuzzy soft set," *Mathematical Problems in Engineering*, vol. 2015, Article ID 582324, pp. 1–9, 2015.

[6] M. A. Jan et al., "Fuzzy soft expert system for classification of lung cancer," *Computers in Biology and Medicine*, vol. 103, pp. 183–197, 2018.

[7] S. P. Anuradha, S. V. Subrahmanyam, "Sugeno fuzzy inference system with adaptive membership functions for breast cancer diagnosis," *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 1, pp. 407–416, 2018.

[8] T. Ahmad, A. Sattar, and R. U. Khan, "A fuzzy soft set approach for medical diagnosis," *Journal of Intelligent & Fuzzy Systems*, vol. 39, no. 2, pp. 1981–1992, 2020.

[9] P. Mandal and D. P. Acharjya, "Fuzzy-soft decision-making system for risk assessment of diabetes," *Applied Soft Computing*, vol. 106, pp. 107316, 2021.

[10] S. Kumar and A. Tripathi, "A hybrid fuzzy inference system for early cancer detection," *Biomedical Signal Processing and Control*, vol. 79, 104017, 2023.

[11] World Health Organization, "Cancer: Key facts," *WHO Report*, Feb. 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer

[12] G. P. Zhang and H. Zhang, "Comparison of Mamdani and Sugeno fuzzy inference systems for classification tasks," *Expert Systems with Applications*, vol. 191, pp. 116243, 2022.

[13] A. Sharma and R. Patel, "Soft set-based clinical decision system for liver disease," *Journal of Healthcare Engineering*, vol. 2020, Article ID 4856431, 2020.

[14] D. Chakraborty et al., "Interpretability in fuzzy rule-based systems: A survey," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 2, pp. 183–197, 2022.

[15] A. Joshi and K. S. R. Anjaneyulu, "A hybrid intelligent model using fuzzy soft set and SVM for early cancer prediction," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 5, pp. 5689–5702, May 2021.

[16] S. H. Almotairi et al., "IoT-enabled fuzzy logic-based early warning system for prostate cancer," *Sensors*, vol. 21, no. 12, pp. 1–16, Jun. 2021.

[17] M. Singh and R. Bhatia, "Hybrid fuzzy-neural model for cancer risk analysis using big data," *IEEE Access*, vol. 9, pp. 74218–74230, 2021.

[18] A. Mahapatra and S. R. Mishra, "Design of a soft computing-based diagnostic model for gastric cancer risk assessment," *Health Information Science and Systems*, vol. 9, no. 1, pp. 1–10, Dec. 2021.

[19] R. Roy and D. Debnath, "A fuzzy soft set model with ensemble learning for intelligent medical diagnostics," *Soft Computing*, vol. 27, pp. 3371–3385, 2023.

[20] Y. Wang and X. Zhang, "Explainable fuzzy deep learning for prostate cancer risk stratification," *Expert Systems with Applications*, vol. 202, 117095, 2022.

[21] N. Dey, A. S. Ashour, and S. Balas, "Fuzzy-based medical expert system for cancer detection," in *Soft Computing Based Medical Image Analysis*, Academic Press, pp. 111–136, 2019.

[22] R. Rani and P. S. Bhattacharjee, "Comparative performance analysis of Mamdani and Sugeno fuzzy models in tumor diagnosis," *Biomedical Engineering Letters*, vol. 10, pp. 345–352, 2020.

[23] A. Rashid et al., "Fuzzy rule-based expert system for colorectal cancer detection using histopathological image features," *IEEE Access*, vol. 10, pp. 56789–56799, 2022.

[24] T. Alshamrani and S. Khan, "Soft set theory in AI-based healthcare applications: A systematic review," *Mathematics*, vol. 11, no. 2, 246, 2023.