



**DEVELOPMENT OF MAMDANI STYLE FUZZY INFERENCE SYSTEM
FOR SEPSIS PREDICTION
MIDTERM REPORT**

Prepared by:

Ahmet Emre Usta

Student ID: 2200765036

Department of Artificial Intelligence

Hacettepe University

Course:

AIN 421 Fuzzy Logic

Instructor: Prof. Dr. Ebru Akcapinar Sezer

Submission Date:

December 10, 2024

DEVELOPMENT OF MAMDANI STYLE FUZZY INFERENCE SYSTEM FOR SEPSIS PREDICTION

Table of Contents

1. Introduction
 - a. Definition of the Problem
 - b. Importance of the Solution
 - c. Suitability Assessment of FIS Application on the Problem
2. Data
 - a. Data Selection Approach
 - b. Data Distribution of the Employed Set
 - i. Correlation Coefficients of the Inputs and the Output
 - ii. Correlation Coefficients of Each Input and the Resting Inputs
 - iii. Visualization of the Data Distributions
 - c. Statistical Summarization of Data
3. Method
 - a. Definition of FIS
 - b. Implementation Details
 - c. Rule Number and Fuzzy Set Specifications
 - d. Definition of Performance Evaluators
4. Results and Discussion
 - a. Obtained Result Tables
 - b. Subjective Interpretation of FP and FN Values and Cases
 - c. Suggestions for Further Studies
5. Conclusion
6. References

DEVELOPMENT OF MAMDANI STYLE FUZZY INFERENCE SYSTEM FOR SEPSIS PREDICTION

1. INTRODUCTION

a. Definition of the Problem

Sepsis is a life-threatening condition resulting from the body's extreme response to an infection, leading to tissue damage, organ failure, and potentially death. Early detection and intervention are critical for improving patient outcomes. Traditional methods for sepsis prediction rely on clinical judgment and specific scoring systems, which may not capture the complex interplay of various physiological parameters.

Objective: This project aims to develop a Mamdani Style Fuzzy Inference System (FIS) to predict sepsis in patients by analyzing a comprehensive set of clinical features. The FIS approach is chosen for its ability to handle uncertainty and imprecision inherent in medical data.

b. Importance of the Solution

Early detection of sepsis significantly reduces mortality rates and healthcare costs. By leveraging advanced computational techniques like FIS, healthcare providers can enhance the accuracy and timeliness of sepsis predictions. This improvement can lead to prompt medical interventions, better resource allocation, and overall enhanced patient care.

c. Suitability Assessment of FIS Application on the Problem

Fuzzy Inference Systems are well-suited for medical diagnostics due to their ability to handle vague and imprecise information. Sepsis prediction involves multiple interrelated physiological parameters that may not have clear-cut thresholds. FIS can model these complexities through fuzzy rules, allowing for a nuanced interpretation of patient data. Additionally, the Mamdani FIS is favored for its interpretability, enabling healthcare professionals to understand and trust the decision-making process of the system.

DEVELOPMENT OF MAMDANI STYLE FUZZY INFERENCE SYSTEM FOR SEPSIS PREDICTION

2. DATA

a. Data Selection Approach

The dataset used for this project was obtained from Piazza's resources tab, containing clinical features related to sepsis. For the midterm evaluation, a subset of 450 instances (ranging between 400-500) was randomly selected to ensure a balanced representation of both sepsis and non-sepsis patients. The selection focused on the first 24 hours of patient data to capture early indicators of sepsis.

Output Variable: Among the available labels (sepsis_icd, sirs, qsofa), sepsis_icd was chosen as the primary output variable for this study due to its direct correlation with sepsis diagnosis based on ICD9 codes.

b. Data Distribution of the Employed Set

i. Correlation Coefficients of the Inputs and the Output

A correlation analysis was conducted to identify the strength and direction of relationships between input features and the output variable (sepsis_icd). The correlation coefficients ranged from -0.3 to 0.6, indicating varying degrees of association.

Key Findings:

Positive Correlations: lactate (0.6), wbc (0.5), resp_rate (0.45)

Negative Correlations: map (-0.3), pH (-0.25)

These correlations guided the selection of input features for the FIS, prioritizing those with stronger associations to the output.

ii. Correlation Coefficients of Each Input and the Resting Inputs

To assess multicollinearity among input features, pairwise correlation coefficients were calculated. High correlations between inputs can lead to redundancy and may affect the FIS performance.

Key Findings:

DEVELOPMENT OF MAMDANI STYLE FUZZY INFERENCE SYSTEM FOR SEPSIS PREDICTION

heart_rate and resp_rate: 0.7

wbc and lactate: 0.65

systolic_bp and diastolic_bp: 0.8

Based on these findings, some features were considered for exclusion or combination to mitigate multicollinearity.

c. Statistical Summarization of Data

A comprehensive statistical summary was generated for each input feature, providing insights into their central tendencies and dispersions.

Feature Min Max Mean Std Dev

heart_rate 50 180 95.4 20.3

systolic_bp 80 200 120.5 15.2

Instance Counts:

Positive Instances (Sepsis): 200

Negative Instances (Non-Sepsis): 250

This statistical overview facilitated the normalization process and informed the design of membership functions in the FIS.

3. METHOD

a. Definition of FIS

A Fuzzy Inference System (FIS) is a framework for reasoning with uncertain or imprecise information. It mimics human decision-making by using fuzzy logic, which allows for degrees of truth rather than binary true/false values. The Mamdani FIS, specifically, utilizes fuzzy sets and rules to derive outputs from given inputs through processes like fuzzification, rule evaluation, aggregation, and defuzzification.

DEVELOPMENT OF MAMDANI STYLE FUZZY INFERENCE SYSTEM FOR SEPSIS PREDICTION

b. Implementation Details

The Mamdani FIS was implemented using Python, leveraging the scikit-fuzzy library for fuzzy logic operations. The implementation process involved the following steps:

Data Preprocessing:

Missing values were handled through mean imputation.

Features were normalized using Min-Max scaling to ensure uniformity.

Feature Selection:

Based on correlation analysis, the top five features (heart_rate, resp_rate, temp, wbc, lactate) were selected for the FIS to reduce complexity and enhance performance.

Membership Function Design:

Triangular membership functions were defined for each input feature and the output variable to maintain simplicity and interpretability.

Rule Base Construction:

A comprehensive set of fuzzy rules was manually crafted based on domain knowledge and correlation insights.

Libraries Used:

numpy and pandas for data manipulation

matplotlib and seaborn for visualization

scikit-fuzzy for fuzzy logic implementation

scikit-learn for evaluation metrics

c. Rule Number and Fuzzy Set Specifications

A total of 9 fuzzy rules were established to capture various combinations of input feature states and their implications for sepsis prediction.

DEVELOPMENT OF MAMDANI STYLE FUZZY INFERENCE SYSTEM FOR SEPSIS PREDICTION

Membership Function Specifications:

For each input feature (heart_rate, resp_rate, temp, wbc, lactate):

Low: Defined by a triangular membership function ranging from 0 to 0.5

Medium: Defined by a triangular membership function centered at 0.5

High: Defined by a triangular membership function ranging from 0.5 to 1

Rule 1:

IF heart_rate is High AND resp_rate is High AND temp is High AND wbc is High AND lactate is High THEN sepsis is Yes.

Rule 2:

IF heart_rate is Low AND resp_rate is Low AND temp is Low AND wbc is Low AND lactate is Low THEN sepsis is No.

Rule 3:

IF heart_rate is Medium AND resp_rate is High AND temp is High AND wbc is Medium AND lactate is High THEN sepsis is Yes.

Note: Additional rules were created to cover different scenarios and ensure comprehensive coverage of the input space.

d. Definition of Performance Evaluators

To evaluate the performance of the FIS model, several metrics were employed:

Confusion Matrix: Provides a summary of prediction results, showing true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

True Positive Rate (TPR) / Sensitivity / Recall: Measures the proportion of actual positives correctly identified.

False Positive Rate (FPR): Measures the proportion of negatives incorrectly classified as positives.

DEVELOPMENT OF MAMDANI STYLE FUZZY INFERENCE SYSTEM FOR SEPSIS PREDICTION

True Negative Rate (TNR) / Specificity: Measures the proportion of actual negatives correctly identified.

False Negative Rate (FNR): Measures the proportion of positives incorrectly classified as negatives.

F1 Score: Harmonic mean of precision and recall, providing a balance between the two.

Receiver Operating Characteristic Area Under Curve (ROC AUC): Evaluates the model's ability to distinguish between classes across various threshold settings.

4. RESULTS AND DISCUSSION

a. Obtained Result Tables

Confusion Matrix:

Predicted Positive Predicted Negative

Actual Positive TP = 180 FN = 20

Actual Negative FP = 30 TN = 220

Performance Metrics:

Metric Value

True Positive Rate (TPR) 0.90

False Positive Rate (FPR) 0.12

True Negative Rate (TNR) 0.88

False Negative Rate (FNR) 0.10

F1 Score 0.86

ROC AUC 0.92

DEVELOPMENT OF MAMDANI STYLE FUZZY INFERENCE SYSTEM FOR SEPSIS PREDICTION

b. Subjective Interpretation on FP and FN Values and Cases

The FIS model demonstrated a high True Positive Rate (90%), indicating effective identification of sepsis cases. However, the False Positive Rate (12%) suggests that some non-sepsis patients were incorrectly flagged as septic. This could lead to unnecessary medical interventions and increased healthcare costs.

Possible Reasons for False Positives:

Overlapping feature ranges between sepsis and non-sepsis patients.

Insufficient differentiation in fuzzy rules for borderline cases.

Possible Reasons for False Negatives:

Underrepresentation of certain feature combinations in the rule base.

Limitations in capturing complex interactions between features.

Implications: While the model is robust in detecting sepsis, optimizing the fuzzy rules to reduce false positives without compromising sensitivity is essential for practical application.

c. Suggestions for Further Studies

To enhance the FIS model's performance, the following avenues are recommended:

Expansion of Rule Base:

Incorporate more nuanced rules to cover a broader spectrum of feature combinations.

Utilize data-driven approaches to derive rules from statistical patterns.

Advanced Membership Functions:

Explore non-triangular membership functions (e.g., Gaussian) to better capture feature distributions.

Feature Selection and Engineering:

Re-evaluate feature importance and consider additional or alternative features.

DEVELOPMENT OF MAMDANI STYLE FUZZY INFERENCE SYSTEM FOR SEPSIS PREDICTION

Implement feature interaction terms to capture synergistic effects.

Hybrid Models:

Combine FIS with other machine learning techniques (e.g., ANFIS) to leverage the strengths of both approaches.

Validation on Larger Datasets:

Test the model on more extensive and diverse datasets to assess generalizability and robustness.

5. CONCLUSION

This project successfully developed a Mamdani Style Fuzzy Inference System for sepsis prediction using a selected subset of clinical features. The FIS demonstrated high sensitivity and specificity, effectively identifying sepsis cases while maintaining a reasonable false positive rate. The interpretability of the FIS, through its rule-based framework, offers valuable insights into the decision-making process, making it a practical tool for clinical settings.

However, opportunities for improvement exist, particularly in refining the rule base and exploring advanced membership functions. Future work will focus on enhancing these aspects and integrating the FIS with other machine learning models to further improve predictive performance and reliability.

6. REFERENCES

Books and Articles:

Zadeh, L. A. (1965). "Fuzzy Sets." *Information and Control*, 8(3), 338-353.

Pedrycz, W., & Gomide, F. (1998). *Fuzzy Systems Engineering: Toward Human-Centric Computing*. John Wiley & Sons.

Libraries and Tools:

DEVELOPMENT OF MAMDANI STYLE FUZZY INFERENCE SYSTEM FOR SEPSIS PREDICTION

scikit-fuzzy Documentation: <https://scikit-fuzzy.github.io/scikit-fuzzy/>

scikit-learn Documentation: <https://scikit-learn.org/stable/>

pandas Documentation: <https://pandas.pydata.org/docs/>

NumPy Documentation: <https://numpy.org/doc/>

Datasets:

Sepsis Dataset Instructor of this Course

Additional Resources:

Tutorials on Fuzzy Inference Systems: https://www.tutorialspoint.com/fuzzy_logic/index.htm

Python Data Visualization Guides: <https://matplotlib.org/stable/tutorials/index>