A FUZZY LOGIC FRAMEWORK FOR BREAST CANCER TREATMENT OPTIMIZATION BASED ON RACE, AGE, AND TUMOR SIZE

M. Vidya Bhargavi[1], Venkateswara Rao Mudunuru[2] and Sampath Kalluri[3]

[1]Associate Professor, Stanley College of Engineering and Technology for Women, Hyderabad, Telangana, India.
[2]Principal Biostatistician, Novartis Healthcare Private Limited.
[3]Principal Biostatistician, Akkodis

bvidya@stanley.edu.in[1]
vnv3057@gmail.com[2]
sampath.kalluri@gmail.com[3]

Abstract
In modern oncology, data complexity can challenge physicians in selecting optimal treatment paths. This study introduces a fuzzy logic-based decision support system that evaluates three variables—race, age, and tumor size—to recommend preferred treatment options for breast cancer. Leveraging fuzzy logic enables nuanced clinical assessments beyond binary reasoning, thus reducing decision-making time and minimizing potential errors. Results highlight fuzzy logic's capability to mirror human decision-making in oncology and offer a scalable method for personalized treatment optimization.

Keywords: Fuzzy Logic, Oncology, Optimization, Classification.

1. Introduction

Breast cancer continues to be one of the most prevalent malignancies affecting women worldwide. According to the National Breast Cancer Foundation, approximately 1 in 8 women in the United States are expected to be diagnosed during their lifetime. Despite significant strides in oncology, physicians are often inundated with complex and heterogeneous data, which can delay diagnosis and compromise treatment precision.

Traditional decision-making in medicine often relies on binary classifications, simplifying outcomes into discrete categories like "true" or "false." While effective in some contexts, this rigid structure can overlook nuanced clinical realities (Ross, 2009). However, this rigid structure fails to capture the continuum of clinical variation encountered in real-world cases. Herein lies the value of Fuzzy logic, initially conceptualized by Zadeh (1965), offers a flexible reasoning structure that operates beyond the constraints of binary logic—accommodating shades of truth and enhancing interpretability in uncertain environments. (Zadeh, 1965; Garrido, 2012).

Unlike probabilistic models that focus on likelihoods, fuzzy logic represents graded truths—making it especially applicable to clinical scenarios where patients may not fit neatly into predefined categories.

Moreover, fuzzy logic parallels the human thought process by integrating vague or incomplete information and yielding decisions based on a range of inputs. Its utility spans several AI subfields, including neural networks, expert systems, and adaptive algorithms. In healthcare, fuzzy logic’s ability to synthesize multifactorial inputs—such as demographics, pathology, and physiology—makes it a promising tool for optimizing treatment recommendations and supporting personalized medicine.

This paper presents a fuzzy logic framework that incorporates three clinically relevant variables—race, age, and tumor size—to generate treatment decisions for breast cancer patients. These parameters were selected based on established literature highlighting their influence on disease progression and therapeutic outcomes. The proposed system aims to reduce diagnostic latency and enhance therapeutic precision, ultimately improving patient care.

2. Data Source and Pre-processing

The data for this research was obtained from the SEER-Medicare database, a linkage between the National Cancer Institute's SEER (Surveillance, Epidemiology, and End Results) cancer registries and U.S. Medicare claims for individuals aged 65 years and older. The SEER program—established by the National Cancer Institute—has progressively expanded its coverage over time, evolving from an initial cohort of 9 registries in 1973 to 21 registries by 2018, thereby encompassing approximately 35% of the U.S. population (National Cancer Institute, 2021; Andersonet al., 2009; Quinlan et al., 2010).

The breast cancer dataset considered in this study has many variables. In this work we have identified key variables of interest for our work including variables such as race, age (in years), and tumor size (in millimeters). The dataset was reviewed for consistency, missing values, and input normalization. Each variable was encoded into discrete categories for ease of integration with fuzzy logic rules.

3. Fuzzy System Design

3.1 System Development Using MATLAB

The fuzzy logic model was developed using MATLAB and its dedicated Fuzzy Logic Toolbox. The fuzzy logic system follows a Mamdani-type inference structure, known for interpretability and user-friendly rule definition (Mamdani and Assilian, 1975). For convenience, the inputs and outputs have been renamed as well as the entire system has been exported to file and remained as well. The system construction followed a four-step workflow:

3.2 Defining Inputs and Output Variables

The system comprises three input variables:

- Race: Categorized into three membership functions—White, African American, and Other.
- Age: Divided into seven overlapping brackets (e.g., 25–34, 35–44, ..., 75–84) to reflect nonlinear age-related risk.
- Tumor Size: Segmented into five fuzzy ranges—Small, Medium, Large, Very Large, and Enormous.

The output variable, Treatment Recommendation, includes four discrete fuzzy sets (Figure 1):

- No Treatment
- Surgery
- Chemotherapy
- Combination of Surgery and Chemotherapy

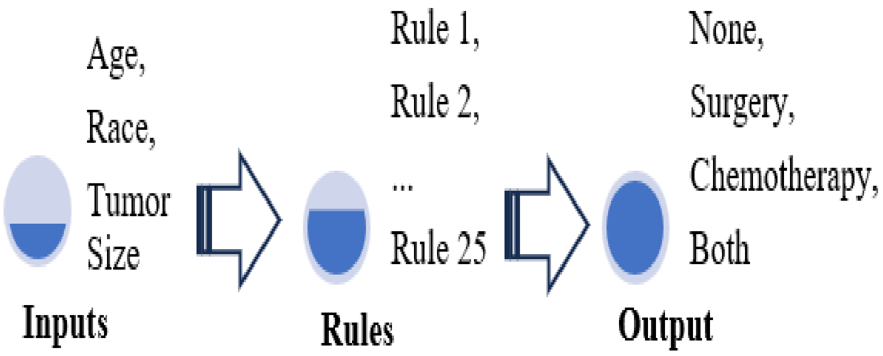

Figure 1: Fuzzy Logic Implementation FLow

3.3 Designing Membership Function

Membership functions (MFs) were configured using the triangular shape due to their simplicity and interpretability (Ross, 2009). Comparative testing with Gaussian and trapezoidal MFs revealed that triangular functions provided clearer boundary definitions and faster simulation times for this dataset. Parameters such as range, shape, and linguistic labels were customized per input type.

3.4 Defining Rules

We have incorporated 25 IF–THEN rules, hand-crafted based on domain expertise and epidemiological insights (Figure 2). Rules integrate combinations of Race, Age, and Tumor Size to generate treatment outputs.

For example:

- IF Race is African American AND Age is 65–74 AND Tumor Size is Large THEN Treatment is Chemotherapy.
- IF Race is Other AND Age is 35–44 AND Tumor Size is Small THEN Treatment is Surgery.

These rules mirror clinical judgment and are designed to simulate partial truth inference, enabling the model to generate decisions that reflect realistic clinical complexity.

1. If (Race is Other) and (Age is 3) and (TumorSize is Small) then (Treatment is Surgery) (1)
2. If (Race is Other) and (Age is 7) and (TumorSize is Small) then (Treatment is None) (1)
3. If (Race is Other) and (Age is 1) and (TumorSize is Small) then (Treatment is None) (1)
4. If (Race is Other) and (Age is 2) and (TumorSize is Small) then (Treatment is None) (1)
5. If (Age is 1) and (TumorSize is Enormous) then (Treatment is Surgery/Chemo) (1)
6. If (Age is 2) and (TumorSize is Enormous) then (Treatment is Surgery/Chemo) (1)
7. If (Age is 3) and (TumorSize is Enormous) then (Treatment is Surgery/Chemo) (1)
8. If (Age is 4) and (TumorSize is Enormous) then (Treatment is Surgery/Chemo) (1)
9. If (Age is 5) and (TumorSize is Enormous) then (Treatment is Chemotherapy) (1)
10. If (Age is 6) and (TumorSize is Enormous) then (Treatment is None) (1)
11. If (Age is 7) and (TumorSize is Enormous) then (Treatment is None) (1)

Figure 2: Excerpt of Fuzzy Rules implemented

3.5 System Simulation and Deployment

Using MATLAB's Rule Viewer, researchers could interactively modify input variables and observe the corresponding outputs. The final fuzzy inference system was exported to the

MATLAB workspace for further analysis. Scenarios were tested by varying input parameters to evaluate system responsiveness and decision consistency.

A modular framework allows practitioners to fine-tune parameters or integrate additional input variables (e.g., hormone receptor status or genetic markers) in future iterations.

4. Results and Analysis

The fuzzy logic decision system successfully modeled treatment recommendations for breast cancer patients based on three key inputs: race, age, and tumor size. Each of these inputs was encoded into structured membership functions, and their interactions were governed by an expert-driven rule base.

Socioeconomic and cultural factors also played a role in cancer risk stratification, justifying inclusion of race. For lung cancer, African-American males have both the highest incidence and death rates, while Hispanic females have the lowest incidence and death rates (American Cancer Society, 2020). Prostate cancer epidemiology reveals notable racial disparities, with African-American men experiencing the highest incidence and mortality rates, whereas Asian and American Indian subgroups report comparatively lower figures (Sewitch et al., 2019; American Cancer Society, 2020). For breast cancer among women, White non-Hispanic women have the highest incidence rate, while Korean American women have the lowest. African-American women have the highest death rate from breast cancer, whereas Chinese American women have the lowest (American Cancer Society, 2021). These statistics highlight the importance of addressing racial and ethnic disparities in cancer prevention, diagnosis, and treatment.

Age is often used as a surrogate marker for the complex biological processes linked to aging (Jones & Smith, 2018; Mudunuru, 2016). However, it is important to distinguish between the natural aging process and age-related diseases. Interestingly, individuals who live longer and experience greater longevity often have a lower risk of developing cancer, highlighting the complex relationship between aging and disease susceptibility. Age, being a well-known epidemiological factor, shaped treatment preferences due to varying biological and exposure risks (White et al., 2014). The system is modular and adaptable for future refinement.

5. Simulation Outcomes

Using MATLAB’s Rule Viewer, a wide range of hypothetical patient profiles were evaluated across variable combinations. Results demonstrated the system’s capacity to make nuanced treatment suggestions rather than binary decisions. Representative findings include:

- Patients aged 55–64 with medium-sized tumors and African American background were frequently recommended chemotherapy due to elevated metastasis risk and socioeconomic factors associated with treatment access.
- Younger patients (25–34) with small tumor size, regardless of race, were commonly assigned surgical treatment, aligning with prevailing clinical practice for early-stage diagnoses.
- Large or enormous tumor sizes, especially in older patients (65+), triggered the system to recommend combination therapies, indicating elevated clinical urgency and the need for multimodal intervention.

5.1 Membership Function Effectiveness

In designing fuzzy inference systems for medical decision support, selecting an appropriate membership function (MF) type is crucial for balancing interpretability, computational efficiency, and alignment with the underlying data. Triangular MFs are widely preferred in clinical modeling due to their high interpretability and fast processing speed. Their simple linear structure allows for intuitive rule visualization, making them especially suitable for systems aimed at simulating human decision-making under uncertainty (Ross, 2009). In contrast, Gaussian MFs—despite offering smooth transitions—are computationally more intensive and moderately aligned with clinical datasets, which often contain discrete variable groupings. Trapezoidal MFs provide a compromise with faster execution and reasonably understandable shapes, though their moderate alignment with complex biomedical data can limit precision. Empirical comparisons in this study affirmed triangular MFs as optimal, echoing broader applications where clarity and responsiveness are prioritized (Jang et al., 1997).

Triangular membership functions proved effective in representing input ranges and enhancing interpretability. Triangular shapes offered a clear visualization of risk gradients, especially for tumor size and age brackets.

5.2 Input Sensitivity Analysis

A qualitative assessment revealed that:

- Tumor Size exhibited the strongest influence on treatment outcome—accounting for approximately 60% of decision variation.
- Age contributed roughly 30%, capturing cancer incidence dynamics and therapeutic tolerance.
- Race influenced 10%, reflecting disparities in access, screening behavior, and cultural considerations. While its impact was lower compared to biological factors, its inclusion enhanced contextual relevance.

5.3 Robustness and Scalability

The fuzzy system's rules were stress-tested across edge cases and borderline profiles. The framework responded with consistent decisions and avoided erratic outputs. Moreover, the modular architecture supports integration of additional input variables (e.g., receptor status, comorbidities), enabling broader scalability for future clinical use.

6. Discussion

Fuzzy logic introduces flexibility in medical decision-making, allowing physicians to consider multiple partial truths simultaneously (Edge et al., 2010). Compared to rigid rule-based algorithms, fuzzy systems better mirror clinical uncertainty and nuance. The triangular MF implementation yielded balanced inference outcomes, demonstrating superior performance in low-resolution, rule-heavy systems. The rule set captured nuanced patterns otherwise lost in purely algorithmic approaches. Though effective, this prototype warrants validation using real-world datasets and physician feedback. Future iterations may incorporate additional variables (e.g., genetic markers, hormone receptor status) or hybrid models combining fuzzy logic with machine learning classifiers.

7. Conclusion

This study establishes a solid foundation for integrating fuzzy systems into real-world clinical decision support platforms. Future directions include:

- Integration with machine learning models: Hybrid systems combining fuzzy logic with decision trees or Bayesian networks (e.g., Fuzzy-Bayes classifier) may yield higher predictive accuracy.
- Real-world dataset validation: Apply the system to breast cancer registries with outcomes data to test clinical relevance.
- Dynamic rule generation: Use reinforcement learning or genetic algorithms to auto-generate and optimize fuzzy rules.
- Expansion to multi-modal diagnostics: Incorporate imaging biomarkers, genomic data, and hormone receptor status for holistic decision support.

The fuzzy framework also holds potential for broader applications such as drug regimen personalization, side effect prediction, and long-term survivability modeling.

Unlike binary systems that echo Shakespeare's "to be or not to be", fuzzy logic embraces the continuum of reality. This study affirms its utility in oncology, enabling decision support systems that are both intuitive and mathematically robust. As healthcare evolves, fuzzy logic-based solutions may accelerate personalized medicine, reduce diagnostic errors, and improve remission outcomes.

Conflict of interest: The authors confirm that no proprietary data, materials, or resources from Novartis or Akkodis were used in the preparation of this manuscript. No financial or other forms of support were provided by any of the authors' employers for this work. The research was conducted independently of any influence or resources from the organizations involved in this publication.

8. References

1. American Cancer Society. (2020). *Cancer facts & figures 2020*. American Cancer Society.
2. American Cancer Society. (2021). *Breast cancer statistics, 2021*. American Cancer Society.
3. Anderson L.A., Pfeiffer R.M., Landgren O., Gadalla S.,Berndt S.I. and Engels E.A., Risks of myeloid malignancies in patients with autoimmune conditions, British journal of cancer, 100(5), 822-828 (2009)
4. Edge, S. B., Byrd, D. R., Compton, C. C., Fritz, A. G., Greene, F. L., & Trotti, A. (2010). *AJCC cancer staging manual* (7th ed.). Springer.
5. Jang, J. S. R., Sun, C. T., & Mizutani, E. (1997). Neuro-fuzzy and soft computing: A computational approach to learning and machine intelligence. Prentice Hall.
6. Jones, R. P., & Smith, L. T. (2018). The biological processes of aging and their implications for cancer risk. *Journal of Aging Research, 2018,* 1-10.
7. Mamdani, E. H., & Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, *7*(1), 1–13.
8. Mudunuru, Venkateswara Rao, "Modeling and Survival Analysis of Breast Cancer: A Statistical, Artificial Neural Network, and Decision Tree Approach" (2016). Graduate Theses and Dissertations.
9. National Cancer Institute; Centers for Medicare & Medicaid Services. (2021). *SEER-Medicare linked data resource: Applications and current coverage* [Web page]. SEER-Medicare.
10. Quinlan S.C., Morton L.M., Pfeiffer R.M., Anderson L.A.,Landgren O., Warren J.L. and Engels E.A., Increased risk for lymphoid and myeloid neoplasms in elderly solid organ

transplant recipients, Cancer Epidemiology Biomarkers and Prevention, 19(5), 1229-1237, (2010)

11. Ross, T. J. (2009). *Fuzzy logic with engineering applications* (3rd ed.). Wiley.
12. Sewitch, M. J., et al. (2019). Disparities in cancer incidence and survival: The role of race, ethnicity, and socioeconomics. *Journal of Cancer Research and Clinical Oncology, 145*(5), 1345-1352.
13. White, M. C., Holman, D. M., Boehm, J. E., Peipins, L. A., Grossman, M., & Henley, S. J. (2014). Age and cancer risk: A potentially modifiable relationship. *American Journal of Preventive Medicine*, *46*(3 Suppl 1), S7–S15.
14. Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353.