An approach to diagnosis of prostate cancer using fuzzy logic

Meena Rawat¹, Pooja Pathak², Pooja Vats³
¹School of Basic and Applied Science, K. R. Mangalam University, Sohna, India
²Department of Mathematics, GLA University, Chaumuhan, India
³Department of Mathematics, School of Basic and Applied Science, K. R. Mangalam University, Sohna, India

ABSTRACT

Early diagnosis of cancers is a major requirement for patients and a complicated job for the oncologist. If it is diagnosed early, it could have made the patient more likely to live. For a few decades, fuzzy logic emerged as an emphatic technique in the identification of diseases like different types of cancers. The recognition of cancer diseases mostly operated with inexactness, inaccuracy, and vagueness. This paper aims to design the fuzzy expert system (FES) and its implementation for the detection of prostate cancer. Specifically, prostate-specific antigen (PSA), prostate volume (PV), age, and percentage free PSA (%FPSA) are used to determine prostate cancer risk (PCR), while PCR serves as an output parameter. Mamdani fuzzy inference method is used to calculate a range of PCR. The system provides a scale of risk of prostate cancer and clears the path for the oncologist to determine whether their patients need a biopsy. This system is fast as it requires minimum calculation and hence comparatively less time which reduces mortality and morbidity and is more reliable than other economic systems and can be frequently used by doctors.

Keywords:
% free prostate-specific antigen
Fuzzy expert system
Fuzzy logic
Prostate cancer
Prostate-specific antigen
Prostate volume

1. INTRODUCTION

In the majority of industrialized nations, prostate cancer is the second most prevalent cause of cancer death for males [1]. Even though the causes are numerous and intricate, they include population growth, aging, and changes in the frequency and distribution of primary cancer incidence and mortality. It is also influenced by a number of variables including a family history of the disease, ethnic origin, and blood levels of the prostate-specific antigen (PSA). The blood PSA level is a crucial factor in the early diagnosis of patients [2]. However, benign prostatic hyperplasia (BPH) and prostate inflammation can both raise the PSA level in blood. It is challenging to distinguish it from BPH for this reason. Prostate biopsy enables the definitive diagnosis of prostate cancer. The PSA, percentage free PSA (%FPSA) test, rectal exam, and transrectal findings assist the clinician to determine whether or not a biopsy is required. However, due to potential consequences and its high cost, those with low cancer risk must avoid this procedure. This allows for the identification of patients with low cancer risk prior to consenting to biopsy. Numerous studies have been conducted on the prognosis or diagnosis of prostate cancer. One of them is fuzzy expert system (FES), a rule-based FES that uses laboratory data (PSA, %FPSA, prostate volume (PV), and patient age) to assess whether a biopsy is necessary. Yuksel et al. [3] proposed a soft expert system using PSA, age, and PV as input variables and PSA as output variables. After comparing their findings to the literature, the devised approach described whether a biopsy was necessary or not. To identify patients at high risk for prostate cancer, Demirtaş and Dalkılıç [4] used an

Journal homepage: http://ijres.iaescore.com
artificial neural network that is related to the consistency of convergence coefficients determined by the fuzzy technique for order preference by similarity to ideal solution (TOPSIS) approach and gives a satisfactory result. Dalkılıç [5] developed a hybrid set model called virtual fuzzy parameterized fuzzy soft set theory, which offers incredibly useful methods for expressing the membership degrees of decision-makers. The notions of point, quasi-coincidence, and mapping were developed for the purpose of building a topological structure on virtual fuzzy parameterized fuzzy soft sets, and some of its distinctive aspects were examined.

Soft sets and fuzzy soft sets in which parameterization tools were re-evaluated and the notions of the pure (fuzzy) soft set were also proposed by Dalkılıç [6] to manage the decision-making processes for uncertainty situations in the most correct manner. Benecchi [7] employed a neuro-fuzzy approach for the diagnosis of prostate cancer of 1,030 patients by using total prostate specific antigen (TPSA) and free prostate-specific antigen (FPSA) and the accuracy given by neuro-fuzzy system was better than that of TPSA and %FPSA. Lee et al. [8] approximately 755 Asian patients with PSA values of 2.0 to 10.0 ng/ml and ages that varied from 40 and 79 years received needle biopsies with TURUS guidance. For this the patients were split into two groups for the diagnosis (144 patients whose PSA range is 2.0-4.0 ng/ml) and (611 patients whose PSA range is 4.1-10.0 ng/ml, n=611), but the authors were unable to achieve any diagnosis result.

It is an extremely difficult task for a decision-maker to determine a value in (0, 1) by using the membership functions because the probability of an error may be very high. Many mathematical models have been constructed to express any situation in the range (0, 1) by taking into consideration hypersoft sets along with fuzzy sets and their derivatives. As a result, Dalkılıç [9] established the ideas of a relational hypersoft membership degree and an inverse relational hypersoft membership degree. The ideas of (NOT) bipolar relational membership degree and (NOT) bipolar relational non-membership degree have been put forth to help establish the values in (0, 1) in a way that is objective and independent of the decision-maker [10].

A fuzzy neural network (FNN) was suggested by Kuo et al. [11] to eliminate the challenge of prostate cancer diagnosis. The optimization version of an artificial immune network was then integrated after cluster analysis membership functions were created for FNN. The particle swarm optimization (PSO) algorithm was created to identify the connection between input and output, and evaluation results revealed that the suggested approach had the capacity to predict prostate cancer and could be readily understood as a result. For the classification of patients who had certain conditions like organ-confined diseases or extra-prostatic disorders, Cosma et al. [12] built a neuro-fuzzy computational model in a study. The primary and secondary gleason biopsy pattern, PSA level, age at the period of detection, and clinical T stage were employed by the authors as input parameters for the diagnosis. The devised approach produced the lowest false positives and the biggest area under the receiver operating characteristic (ROC) curve.

In a study, neuro-fuzzy classification (NEFCLASS), an innovative approach for the diagnosis of prostate cancer was developed by Keles et al. [13] and was based on neuro-fuzzy categorization. This method has some unique features, including batch learning, automatic cross-validation, the ability to manage missing values, and the automatic determination of rule base size, all of which improve its interpretability. Seker et al. [14] introduced a fuzzy K-nearest neighbour classifier (FKNN) which gives a certain degree for decisions regarding diagnosis in the evaluation of breasts and prostate cancer and compares the outcome with logistic regression and multilayer back propagation neural network as an artificial neural network tool. For predicting these markers, FKNN approaches were the most accurate. FKNN’s breast cancer node predictive rate is 88%, and LR’s prostate cancer predictive rate is 68.3%. In order to build their FES, Kar and Majumder [15] used four linguistic variables for each of the three input factors (PSA, PV, and age). When a neural network is trained using the Levenberg-Marquardt training technique, it functions satisfactorily and achieves 100% accuracy.

Expert systems have been created in a variety of fields. Fuzzy logic has a crucial role in medical for the improvement of decision-making in radiation therapy, detection of breast cancer, lung cancer, and prostate cancer, to differentiate of benign skin lesions from malignant melanoma, an MRI-based method for calculating the volume of brain tissue, to assist doctors in making a rapid and effective decision regarding the dose of medicine for the treatment of 200 dialysis patients [16]–[21]. When doctors begin treating a patient, they analyze all medical records and use their academic knowledge, personal experience, and intellectual power to determine the origin of the ailment. As a result, the purpose of the fuzzy intelligent system is to control the behavior of doctors by providing them with a conference. Fuzzy logic addresses several aspects of FESs, and its applications in the medical domain are vast [22].

Fuzzy logic is now widely used by doctors and engineers to tackle difficulties in a variety of disciplines, including agriculture, banking and economics, electronics, and other real-world issues. Previous research has shown that detecting prostate cancer using solely image processing and ultrasonography is difficult. For this, we created a knowledge-based FES that uses the patient’s laboratory and other data and mimics the expert’s advice. If it is found early, the patient can be thoroughly treated, increasing the patient’s chances of survival. On the other hand, biopsy for cancer detection has the potential to spread to other organs. As a result, the biopsy process is unappealing and unacceptably painful. We developed a fuzzy rule-based
A technique that incorporates PSA, PV, age, and %FPSA as input elements and prostate cancer risk (PCR) as an output component to determine how much biopsy is required [23].

2. RESEARCH METHOD

2.1. Fuzzy expert system

FES stands for fuzzy expert system, which substitutes fuzzy logic rather than Boolean logic. A FES is a type of artificial intelligent (AI) that uses membership functions and a set of predetermined rules to assess a batch of data. A FES’s sharp inputs are given fuzziness by using the right membership functions. All input variables are supplied to a specific inference technique after their membership functions have been defined. The Mamdani (max-min) inference approach, which is the most well-liked in the literature, was applied in this case. The FES rules created here follow an IF-THEN format. To produce crisp output, the fuzzy output is defuzzified using a variety of approaches. In our FES, the centroid approach is used for defuzzification. Figure 1 depicts the basic layout of a FES.

![Figure 1. General structure of FES](image)

The output of the Mamdani type inference method is fuzzy sets. This expert system is fuzzy in nature since the rules in the expert system incorporate fuzziness and the parameters can be fuzzified [24]. Defuzzification is a crucial concept in the field of fuzzy logic, primarily used to convert the fuzzy output of a fuzzy inference system into a crisp, understandable value. It plays a pivotal role in making fuzzy logic applicable to real-world problems, as it transforms the imprecise, fuzzy sets and membership degrees into concrete, numerical values that can be readily used for decision-making or control purposes.

2.2. Dataset

The dataset we have considered for input variable from the earlier article contributed by Saritas et al. [25]. As per researchers Saritas et al. [25], this data set is a collection of real clinical data of 119 patients in Ankara University Medicine Faculty in 2005. Age (year), PV (ml), PSA (ng/ml), FPSA (%), and PCR (%) are the units used for input and PCR (%) for output parameters. Here we developed FES and calculated the risk of PCR by using the dataset of 119 patients with their PSA level, PV, age and %FPSA and the result of biopsies in the Saritas et al. [25].

3. PROPOSED METHOD

In the ever-evolving landscape of data-driven solutions, the need for efficient algorithms is paramount. In response to this demand, we propose an algorithm designed to tackle a specific set of challenges. The proposed algorithm builds upon existing methodologies, yet introduces innovative elements that promise to enhance performance and versatility in various domains. This algorithm represents a significant step forward in addressing critical problems, offering a fresh perspective, and potential solutions to complex computational tasks. The proposed algorithm is clearly shown in Algorithm 1.
Algorithm 1. Proposed FES approach
Proposed Algorithm
Step 1. Begin
Step 2. Preprocessing the data
Step 3. Take fuzzy sets for input and output variables for the fuzzification
Step 4. Decide a range for each fuzzy set
Step 5. Formation of membership function for input and output variables
Step 6. Formation of fuzzy knowledge rule base
Step 7. Select the methodology for defuzzification
Step 8. Generate a fuzzy expert system model
Step 9. Apply the patient’s clinical data to the model
Step 10. Observe all the output values.
Step 11. Construct the performance matrix for the clinical data
Step 12. Calculate the classification accuracy, True positive and true negative of the data sets
Step 13. End

3.1. Input variables array
Most of the earlier researcher’s contributions we observed that using ‘triangular and trapezoidal functions enhanced the performances. Researchers like Zhao and Bose [26] compared the response of the system with various membership functions and conveyed that the triangular membership function is superior to any other membership functions. In light of the data set selected for our proposal, we have chosen the following input variables: ‘Age’, ‘PSA’, ‘PV’, ‘%FPSA’.

3.1.1. Age
The age of a person has a big impact on how much prostate cancer costs [27]. Prostate cancer is most commonly detected in men over the age of 65. The number differs from one person to the next. For age, four fuzzy sets are used: “Very Young”, “Young”, “Middle”, and “Old”. After that, trapezoidal membership functions are used to interpret all four fuzzy sets. Figure 2 shows the graph of the membership function of age.

3.1.2. Prostate-specific antigen
PSA is a protein produced by the prostate gland and found primarily in the sperm, however, a small quantity can also be found in the blood [28]. When a person gets older or their prostate gland grows larger, the amount of PSA in their blood increases somewhat. A high PSA level indicates that you may have a problem with your prostate, but it does not always mean that it is cancer. A healthy prostate, according to specialists, produces less PSA in the blood than a malignant one. As a result, an increase in PSA could be a sign of prostate cancer. For PSA five fuzzy sets: “Very Low”, “Low”, “Middle”, “High”, and “Very High” are used, and for first and fifth fuzzy sets are interpreted by trapezoidal membership function while for the remaining we used triangular membership function. Figure 3 shows the graph of the membership function of age.

3.1.3. Prostate volume
The prostate of a healthy guy is about the size of a chestnut, and its role is to create a small amount of seminal fluid, which combines with sperm to form semen [29]. The prostate gland begins to expand at the age of 40, but the pattern varies from person to person. It's critical for the early detection of prostate cancer. Total PV is calculated using the elliptical formula, TPV, which is defined as π/6 W L H, where W denotes prostate width, L denotes prostate length, and H denotes prostate height. PV uses four fuzzy sets: “Small”, “Medium”, “Large”, and “Very Large”. The trapezoidal membership function is used to interpret the first and fourth fuzzy sets, whereas the triangle membership function is used to understand the second and third fuzzy sets. The graph of PV’s membership function is displayed in Figure 4.

3.1.4. Percentage free prostate-specific antigen
PSA testing provides total PSA, which includes both bound and unbound PSA levels, whereas %FPSA provides the ratio of unbound to bound PSA levels by calculating (free/total PSA)×100% [30], [31]. On average, a FPSA test can cut the number of needless biopsies by 20%. The higher the PSA level and the lower the %FPSA level, the greater the risk of prostate cancer. For %FPSA, four fuzzy sets are used: “Low”, “Medium”, “High”, and “Very High” with the first and fourth fuzzy sets interpreted using a trapezoidal membership function, and the second and third fuzzy sets interpreted using a triangle membership function. The graph of %FPSA's membership function is displayed in Figure 5.
3.2. Output variables array: prostate cancer risk

The fuzzy sets are classified as “Very Low”, “Low”, “Middle”, “High”, and “Very High” and the first and fourth fuzzy sets are interpreted using a trapezoidal membership function, while the remainder are interpreted using a triangle membership function. The created technology can diagnose and treat a variety of prostate cancers in patients. Figure 6 depicts the PCR membership function graph.
### 3.3. Formation of fuzzy knowledge rule base and defuzzification

We utilize four input variables, age, PV, PSA, and %FPSA, as inputs, and interpret these variables using four, four, five, and four membership functions, accordingly. Total 255 rules are setup to calculate PCR which is expressed by five membership functions. Table 1 is a few fuzzy rules we developed.

<table>
<thead>
<tr>
<th>Rule No.</th>
<th>Age</th>
<th>PV</th>
<th>PSA</th>
<th>%FPSA</th>
<th>PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Middle</td>
<td>Small</td>
<td>V. Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>53</td>
<td>Middle</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>92</td>
<td>Middle</td>
<td>V. Large</td>
<td>Low</td>
<td>High</td>
<td>Middle</td>
</tr>
<tr>
<td>150</td>
<td>Old</td>
<td>Small</td>
<td>V. High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>191</td>
<td>Old</td>
<td>Large</td>
<td>Middle</td>
<td>Medium</td>
<td>Middle</td>
</tr>
<tr>
<td>200</td>
<td>Old</td>
<td>Large</td>
<td>High</td>
<td>V. High</td>
<td>V. High</td>
</tr>
<tr>
<td>230</td>
<td>Young</td>
<td>Small</td>
<td>Low</td>
<td>V. High</td>
<td>V. Low</td>
</tr>
</tbody>
</table>

### 3.4. Defuzzification by using Mamdani fuzzy inference system

Defuzzification is a technique for extracting a single quantity from a set of fuzzy numbers. It converts the output of fuzzy inference into a crisp output. The crisp value of PCR in this FES is derived using the centroid method of defuzzification, which may be calculated using (1).

$$D^* = \frac{\int_{a}^{b} \mu_c(x) d x}{\int_{a}^{b} \mu_c(x) dx}$$

The mass centroid notion gave rise to the centroid defuzzification technique used in fuzzy systems to determine the centroid of a homogeneous piece of material. If a mass is held at its centroid point, it will balance correctly [32], [33]. This is the center of the mass. Thus, if we are given a set of input data for prostate cancer detection, such as age=67, PSA=9.32, PV=36, %FPSA=20.37, we must first choose the appropriate fuzzy sets for each input variable and calculate the necessary degree of membership function. Assume that this set of inputs adheres to some (k) set of rules. The truth degree I of the ith applicable rule is determined after determining the minimum of related membership values for each input variable. By taking the maximum of all we can determine the PCR membership value. In this case, the Mamdani fuzzy inference technique is applied. The PCR crisp value is calculated using the centroid method. The following are the phases for the mentioned data:

- Age=67, \( \mu_{\text{middle age}}(67)=0.8 \), and \( \mu_{\text{old}}(67)=0.2 \),
- PSA=9.32, \( \mu_{\text{middle}}(9.32)=0.795 \), and \( \mu_{\text{high}}(9.32)=0.164 \),
- PV=36, \( \mu_{\text{small}}(36)=0.2 \), and \( \mu_{\text{medium}}(36)=0.2 \),
- %FPSA=20.37, \( \mu_{\text{medium}}(20.37)=0.61 \), and \( \mu_{\text{high}}(20.37)=0.0491 \).

According to the given input data, sixteen rules will be fired as following:

- \( a_1 = \min(0.8,0.2,0.795,0.61)=0.2 \),
- \( a_{15} = \min(0.8,0.2,0.795,0.0491)=0.0491 \),
- \( a_{21} = \min(0.8,0.2,0.164,0.61)=0.164 \),
- \( a_{53} = \min(0.8,0.2,0.164,0.0491)=0.0491 \),
- \( a_{44} = \min(0.8,0.2,0.795,0.61)=0.2 \),
- \( a_{46} = \min(0.8,0.2,0.795,0.0491)=0.0491 \),
- \( a_{57} = \min(0.8,0.2,0.164,0.61)=0.164 \).
Now, we obtain membership function from Mamdani max-min inference as:

$$\alpha = \max(\alpha_{13}, \alpha_{15}, \alpha_{21}, \alpha_{23}, \alpha_{44}, \alpha_{46}, \alpha_{52}, \alpha_{54}, \alpha_{131}, \alpha_{139}, \alpha_{141}, \alpha_{163}, \alpha_{165}, \alpha_{171}, \alpha_{17})$$

$$= \max(0.0, 0.0491, 0.164, 0.0491, 0.164, 0.0491, 0.164, 0.0491, 0.2, 0.0491, 0.0491, 0.2, 0.0491, 0.164, 0.164, 0.0491) = 0.2.$$  

After applying the centroid method, we obtain PCR=56% which is slightly greater than our cut-off 50%. In this situation, we will advise the patient to go for a biopsy. Figure 7 shows the calculation of PCR for age 67 years, PSA=9.32 ng/ml, PV=36 ml and %FPSA=20.4. Thus, applying the developed FES to the data of 119 patients with their PSA level, PV, age, and FPSA to calculate the risk of PCR and then compare the results with the result of biopsies given in the paper [25]. In light of the FIS result indicating a 56% probability of prostate cancer, it is prudent to recommend a biopsy for a comprehensive evaluation. Our established cutoff is 50% underscores the importance of further investigation, as early detection and timely action can significantly impact treatment outcomes.

![Figure 7. Calculation of PCR for age=67 years, PSA=9.32 ng/ml, PV=36 ml, and %FPSA=20.4](image)

4. RESULTS AND DISCUSSIONS

We examine the results of biopsies taken from 119 patients who have acquired FES. The majority of FES for PCR literature focuses on three inputs (PSA, age, PV). However, the %FPSA with PSA plays a critical role in the identification of PC, reducing unnecessary biopsy, and yielding significant outcomes. That is why we created our fuzzy-rule knowledge-based expert system, which uses four input factors and one output parameter to predict the risk of prostate cancer. If our FES indicates that the PCR range equals or exceeds 50%, the patient should be encouraged to get a biopsy. The findings of our FES are shown in Table 2.

Out of 119 patients, 61 had positive biopsy results, while the remaining 58 had negative biopsy results. Our FES has a true positive detection rate of 77.05%. Out of 61 positive patients, our FES properly identified 47 positive biopsy results and accurately identified 30 negative biopsy results out of 58 patients. The proportion of proper detection after utilizing our developed FES is significantly higher than that of other existing methods in the literature. Using fuzzy rules on the same data, Saritas got 64.71% correct detection, 62.18% with an online calculator, and 60.50% with the ratio FPSA/PSA for PCR. So, while the addition of %FPSA percent increased the number of rules, it also improved the accuracy of the findings.
5. CONCLUSION

The true detection percentage for positive biopsy is great, although it varies slightly for negative biopsy cases. Our FES is not conclusive in determining whether a patient has cancer or not. However, it provides a percentage of the likelihood of prostate cancer and aids the doctor in determining whether or not a patient should get a biopsy. Because diverse immune modulation behaviors range from person to person, demographic variations from area to area, family history, height, eating habits, style of life, and many more hidden activities are all factors. As a result, moderation in determining cut-off values must be developed in accordance with the FES’s terms and conditions. We can improve our FES by incorporating other artificial techniques.

ACKNOWLEDGEMENTS

Authors are very thankful for the outstanding support and moral help for this research work by Dr. Neeraj Kumar, who has huge knowledge in the same research domain.

REFERENCES


Table 2. Confusion matrix

<table>
<thead>
<tr>
<th>Performance metric</th>
<th>Actual positive (61)</th>
<th>Actual negative (58)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted positive</td>
<td>47</td>
<td>28</td>
</tr>
<tr>
<td>Predicted negative</td>
<td>14</td>
<td>30</td>
</tr>
<tr>
<td>Accuracy</td>
<td>64.71</td>
<td></td>
</tr>
<tr>
<td>True positive</td>
<td>77.05</td>
<td></td>
</tr>
<tr>
<td>True negative</td>
<td>51.72</td>
<td></td>
</tr>
</tbody>
</table>

An approach to diagnosis of prostate cancer using fuzzy logic (Meena Rawat)


**BIOGRAPHIES OF AUTHORS**

**Meena Rawat** is a research scholar in the School of Basic and Applied Science, K. R. Mangalam University, Gurugram, Haryana. She has M.Sc. in applied mathematics from Bhim Rao Ambedkar University and M.Phil. in mathematics from Madurai Kamaraj University. She has 16 years of teaching experience. Currently, she is working on the diagnosis of different types of cancers using soft computing techniques. She can be contacted at email: rawatmeenams@gmail.com.

**Pooja Pathak** is associate professor, IAH, GLA University, Mathura (India) earned her Ph.D. from BRA Agra University in the Year 2010. She has published more than 40 papers and articles in reputed journals and conferences. Her area of research includes neuro-fuzzy-genetic, and soft computing. She has 21 years of experience along with teaching and research exposure. She has more than twenty years of experience in teaching graduate and post-graduate and adaptive by nature new technology in the field of computer science. She can be contacted at email: pooja.pathak@glac.ac.in.

**Pooja Vats** is assistant professor, K. R. Mangalam University, Gurugram (India) completed her Ph.D. from K. R. Mangalam University. She has 14 years of teaching experience. Throughout the course of her career, she has taught a wide variety of courses at both UG and PG levels. She is always dedicated to teaching and also interested in the field of research and innovation. She has published eight research papers and one book chapter in reputed journals. She has been actively involved in the curriculum design for the students of mathematics across various courses. She can be contacted at email: pooja.vats@krmangalam.edu.in.