AITA

2024 Volume 4 Number 2

ARTIFICIAL INTELLIGENCE THEORY AND APPLICATIONS

https://aitajournal.com

info@aitajournal.com

ISSN: 2757-9778

ISBN: 978-605-69730-2-4

ARTIFICIAL INTELLIGENCE
THEORY AND APPLICATIONS

Volume 4 – Issue 2

01.10.2024

e-ISSN: 2757-9778

Publisher: İzmir Bakırçay University

www.aitajournal.com

https://dergipark.org.tr/en/pub/aita

Journal Board

AITA

Artificial Intelligence Theory and Applications

Editorial Board

International Advisory Board

International Indexes

ACADEMIC RESOURCE INDEX: RESEARCH BIBLE

INDEX COPERNICUS INTERNATIONAL

SOBIAD CITATION INDEX

ASIAN SCIENCE CITATION INDEX

INTERNATIONAL INSTITUTE OF ORGANIZED RESEARCH

EUROPUB INDEX

About Journal

Artificial Intelligence Theory and Applications (AITA) provides coverage of the most significant work on principles of artificial intelligence, broadly interpreted. The scope of research we cover encompasses contributions of lasting value to any area of artificial intelligence. To be accepted, a paper must be judged to be truly outstanding in its field. AITA is interested in work in core artificial intelligence and at the boundaries, both the boundaries of sub-disciplines of artificial intelligence and the boundaries between artificial intelligence and other fields.

Scope

The best indicator of the scope of the journal is provided by the areas covered by its Editorial Board in theoretical (artificial intelligence and computing methodologies) and practical (artificial intelligence applications and applied computing) ways. These areas change from time to time, as the field evolves.

Year : 2024
Volume : 4
Issue : 2

ARTIFICIAL INTELLIGENCE THEORY AND APPLICATIONS

Table of Content **Page**

AITA
ARTIFICIAL INTELLIGENCE THEORY AND APPLICATIONS

■ RESEARCH ARTICLE

Prioritizing Digital Health: Key Municipal Services Identified Through Fuzzy Methods

Aleyna Erdoğan[a], **Gizem Turcan** [b†], **Onur Doğan**[b], **Erman Coşkun**[b]

[a] Department of Industrial Engineering, Pamukkale Universtiy, Denizli, Türkiye
[b] Department of Management Information Systems, İzmir Bakırçay University, İzmir, Türkiye
† gizem.turcan@bakircay.edu.tr, corresponding author

RECEIVED APRIL 24, 2024
ACCEPTED JULY 16, 2024

CITATION Erdoğan, A., Turcan, G., Doğan, O., & Coşkun E. (2024). Prioritizing digital health: key municipal services identified through fuzzy methods. *Artificial Intelligence Theory and Applications, 4(2*), 76-106.

Abstract

The integration of digital technologies into healthcare systems within municipalities has elicited a transformative change in health service delivery. This paper explores the importance of the digitalization of health services in municipalities and represents the most important services by employing fuzzy methods. The research evaluates the importance of digital transformation of several health services in municipalities by examining existing literature and employing a combination of qualitative and quantitative methods, including the Pythagorean Fuzzy CRITIC (PF-CRITIC) and Interval-Valued Pythagorean Fuzzy WASPAS (IVPF-WASPAS) methods. Key findings highlight that mobile health services and medical center services are the two most important municipal health activities regarding digital transformation. Additionally, we employed sensitivity analysis to assess the stability and reliability of the methods, thereby conducting a detailed analysis of the decision-making process. Through evidence-based strategies, municipalities can harness the power of digitalization to develop patient-centered, efficient, and responsive healthcare services. Therefore, this study contributes to a more inclusive approach to digitalization in healthcare, aiming to obtain the opinions of individuals who have experience with health activities in municipalities.

Keywords: healthcare services, municipalities, fuzzy, digitalization

---

1. Introduction

The convergence of digital technologies and healthcare has significantly transformed the delivery of health services in municipalities in recent years. This transition towards digitalization represents a transformative journey, promising enhanced efficiency, accessibility, and quality of healthcare provision. The integration of digital tools into health services, from electronic health records to telemedicine platforms and mobile health applications, holds immense potential to revolutionize the access, delivery, and experience of care for both patients and providers.

The emergence of digitalization faces both opportunities and challenges within the healthcare workforce. While digital technologies have the potential to streamline workflows, improve communication, and empower healthcare professionals, there is an

Artificial Intelligence Theory and Applications, ISSN: 2757-9778. ISBN: 978-605-69730-2-4 © 2024 İzmir Bakırçay University

obvious doubt about digital health solutions in the workforce. Addressing the concerns surrounding digital literacy, privacy, and data security is crucial to foster a culture of acceptance and readiness for digital transformation within the healthcare sector.

The digitalization of healthcare services in municipalities represents a comprehensive transformation that includes the integration of digital technologies and information systems into various aspects of healthcare. This paradigm shift varies from the digitization of medical records to include telemedicine platforms, mobile health applications, wearable devices, remote monitoring systems, and advanced analytics. Digitalization aims to improve communication between healthcare providers and individuals, improve access to care, optimize resource allocation, and improve health outcomes by streamlining processes. Accordingly, municipalities' adoption of digitalization offers the potential to improve healthcare services, making them more patient-focused, efficient, and responsive to the evolving needs of society.

This research has a combination of qualitative and quantitative approaches. The first phase comprises a literature review to synthesize existing knowledge for understanding the impact of digitalization on health services within municipalities. Subsequently, we employ the Pythagorean Fuzzy CRITIC (PF-CRITIC) method and the Interval-Valued Pythagorean Fuzzy WASPAS (IVPF-WASPAS) method to analyze data and evaluate the various aspects of digitalization in health service delivery. This integrated approach aims to provide insights for decision-making and technology development to improve health services by evaluating the digital transformation's impact on various health services in municipalities. Additionally, we used sensitivity analysis to evaluate the reliability of the model.

In light of these considerations, this research endeavors to explore the multifaceted impact of digitalization on health services within municipalities. Accordingly, by focusing on the needs and experiences of society, this paper aims to contribute to a more inclusive approach to digitalization in healthcare.

2. Literature Review

The digitalization of healthcare services in municipalities is a growing subject, focusing particularly on the impact of digitalization on public service delivery for socially disadvantaged individuals. Buchert et al. [1] emphasize the lack of empirical research examining the effects of digitalization on public health and social welfare services from the perspective of socially disadvantaged individuals and emphasize the need for more comprehensive studies in this field. So, Schou & Pors [2] discuss the shift towards self-service solutions in welfare services due to digitalization, which places the responsibility on citizens to actively seek services previously managed by professionals, raising concerns about the potential exclusion of disadvantaged individuals.

In the public sector domain, Lloyd & Payne [3] address the use of digitalization as a cost-effective method for delivering better care quality and more client-focused services, reflecting the ongoing efforts to leverage digitalization for improved public health services. Additionally, Collington [4] highlights the emergence of public sector digitalization strategies with the goal of improving services and enhancing efficiency, indicating a broader trend towards digital transformation in public service delivery.

The impact of digitalization on health care professionals and citizens is also a significant area of concern. Tiainen et al. [5] point out that digitalization poses challenges not only for health care and social welfare professionals but also for citizens, highlighting the need for comprehensive strategies to address the implications of digitalization in these sectors. Moreover, Baumgartner et al. [6] note a questioning attitude towards digital health among medical students, indicating the importance of addressing perceptions and preparedness for digitalization in the health care sector.

The literature on the digitalization of health services in municipalities is extensive and diverse, covering various aspects of digital transformation in healthcare, public health, and social welfare services. Gopal et al. [7] address the importance of digital transformation in healthcare, highlighting the integration of technologies like the Internet of Things, advanced analytics, Machine Learning, and Artificial Intelligence as key components to address challenges in healthcare. Scarano & Colfer [8] discuss the review of automated possibilities in linking active labor market policies to digitalization, considering the potential impact on employment and public services. Holm et al. [9] provide insights into the allocation of home care services by municipalities in Norway, indicating potential equity issues in the allocation system. Moreover, Collington [4] examines how digitization affects the capacity reduction of the public sector, emphasizing the need for more study on how governments might use technological advancement for the benefit of their population while keeping themselves functional during the process.

These studies provide a comprehensive overview of the multiple impacts of digitalization on health services in municipalities, addressing technological integration, service allocation, ethical considerations, and the broader implications for public sector capacity.

The implications of digitalization in the health services of municipalities have significant effects on various aspects of service delivery. Buchert et al. [1] highlight the reinforcement of social exclusion through the digitalization of public health and social welfare services, particularly for disadvantaged individuals. This underscores the need for comprehensive strategies to address the potential negative impact of digitalization on vulnerable populations. Additionally, Schou & Pors [2] emphasize the qualitative study of exclusion in digitalized welfare, shedding light on the impact of digitalization on welfare institutions and professional practices, particularly in the context of disadvantaged individuals. These findings underscore the complex interplay between digitalization, public sector capacity, and citizen welfare, emphasizing the need for careful consideration of the implications of digitalization in health services.

Additionally, Holm et al. [9] provide insights into the allocation of home care services by municipalities, indicating potential fairness issues in the allocation system. This highlights the need for equitable and transparent digitalized processes for service allocation to ensure fair access to health services. Furthermore, Shava & Vyas-Doorgapersad [10] highlight the need for comprehensive digital infrastructure to support effective service delivery, pointing out that municipalities are unable to foster digital innovations to improve public service delivery due to a lack of digital skills, infrastructure, accessibility, and connectivity.

The integration of qualitative and quantitative methods in studying the impact of digitalization on health services is well-supported by existing literature. O'cathain [24] used a mixed methods approach to evaluate the impact of health information systems in UK. Their use of both qualitative interviews and quantitative data analysis provided a comprehensive understanding of the system's effectiveness, similar to our approach.

Moreover, the application of multi-criteria decision-making (MCDM) methods, such as PF-CRITIC and IVPF-WASPAS, has been validated in various fields, including health services. Haktanir and Kahraman [25] utilized the CRITIC method to assess the performance of healthcare providers, and Wang et al. [26] applied PF-CRITIC method to select suppliers, while Gedikli and Cayir Ervural [27] applied the IVPF-WASPAS method to prioritize COVID-19 vaccine alternatives. These studies demonstrate the robustness and applicability of these methods, supporting our choice of methodology for evaluating digitalization impacts.

The literature underscores the need for more empirical research on the effects of digitalization on public health and social welfare services. There is also a growing focus on the challenges and implications of digitalization for health care professionals and citizens, highlighting the need for comprehensive strategies to address the impact of digitalization on health services in municipalities.

3. Preliminaries

3.1. Pythagorean Fuzzy Sets

Yager [11] proposed Pythagorean Fuzzy Sets (PFS) based on the logic of Intuitionistic Fuzzy Sets (IFS), which was developed by Atanassov [11], in 2013. In IFS, the sum of the degrees of membership (μ) and non-membership (υ) of an element in a set is in the range [0,1]. In the PFS, however, the sum of the squares of the degrees of membership and non-membership of an element cannot exceed 1. The PFS, as an extension of the IFS, allow experts to make evaluations on a wider scale [12] [13]

For example, a decision maker may determine the membership degree of an alternative to be √3/2 and the non-membership degree to be 1/2. In this case, since the sum of the membership and non-membership degrees exceeds 1, the use of IFS is not appropriate. However, since the condition $0 \leq (\frac{\sqrt{3}}{2})^2 + (\frac{1}{2})^2 \leq 1$ is satisfied, PFS can be used. In this regard, instead of asking decision makers to adjust their decisions to fit within the limits of IFS, PFS can be used. It is claimed that PFS have more capability than IFS in modeling uncertainty for decision-making problems [13].

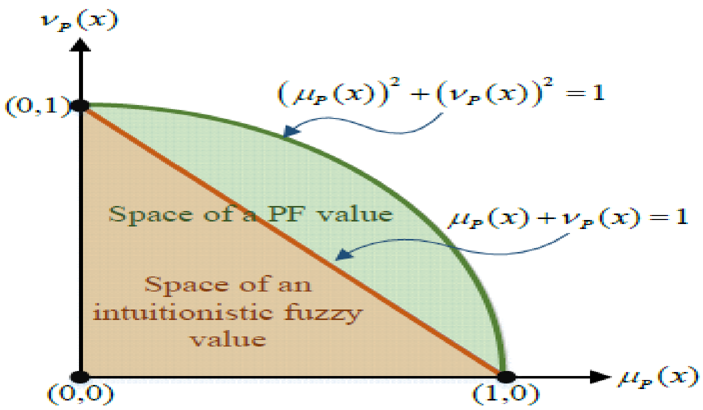


Figure 1. Comparison of PFS and IFS [12]

The comparison between IFS and PFS is provided in Figure 1. According to this figure, it is observed that PFS encompass IFS. PFS differ from IFS in that PFS allow the sum of membership and non-membership degrees to exceed 1, but the sum of their squares cannot exceed 1 [12].

Definition: Let X be the universal set and P be the Pythagorean fuzzy set object of this universal set. P object is defined as seen in Equation (1) [14]:

$$P = \left\{< x, P\left(\mu_p(x), v_p(x)\right) > \mid x \in \mathrm{X}\right\} \tag{1}$$

Here, $\mu_p(x)\colon X \mapsto$ [0,1] represents the membership degree, and $v_p(x)\colon X \mapsto$ [0,1] represents the non-membership degree.

The sum of the squares of the membership and non-membership degrees of an element x in the universal set X, belonging to the subset P, as seen in Equation (2), does not exceed 1 [14].

$$0 \leq \mu_p(x)^2 + v_p(x)^2 \leq 1 \tag{2}$$

3.1.1. Interval-Valued Pythagorean Fuzzy Sets

Peng and Yang [15] expressed fuzzy sets as Interval-Valued. Accordingly, membership and non-membership degrees are defined within lower and upper bound intervals. These sets are named as Interval-Valued Pythagorean Fuzzy Sets (IVPFS).

An Interval-Valued Pythagorean Fuzzy Set p̃ in the universe X is defined as follows: if x is an element, μ represents the membership degree, υ represents the non-membership degree, and L and U represent the lower and upper bounds of these degrees, respectively, as shown in Equation (3). The sum of the squares of membership and non-membership degrees does not exceed 1, as illustrated in Equation (4) [15] [16].

$$\tilde{p} = \left\{\left(x, \left[\,\mu_{\tilde{p}}^{L}(x), \mu_{\tilde{p}}^{U}(x)\right], \left[\,\upsilon_{\tilde{p}}^{L}(x), \upsilon_{\tilde{p}}^{U}(x)\right]\right); x \in X\right\} \tag{3}$$

$$0 \leq \left(\mu_{\tilde{p}}(x)\right)^2 + \left(\upsilon_{\tilde{p}}(x)\right)^2 \leq 1 \tag{4}$$

4. Methodology

4.1. The Pythagorean Fuzzy CRITIC (PF-CRITIC) Method

The Pythagorean Fuzzy CRITIC (PF-CRITIC) method, introduced into the literature by Peng, Zhang, and Luo in 2020 [17], is an adaptation of the classical CRITIC method to the Pythagorean fuzzy numbers. The process flow diagram of the PF-CRITIC method is modeled in Figure 2.

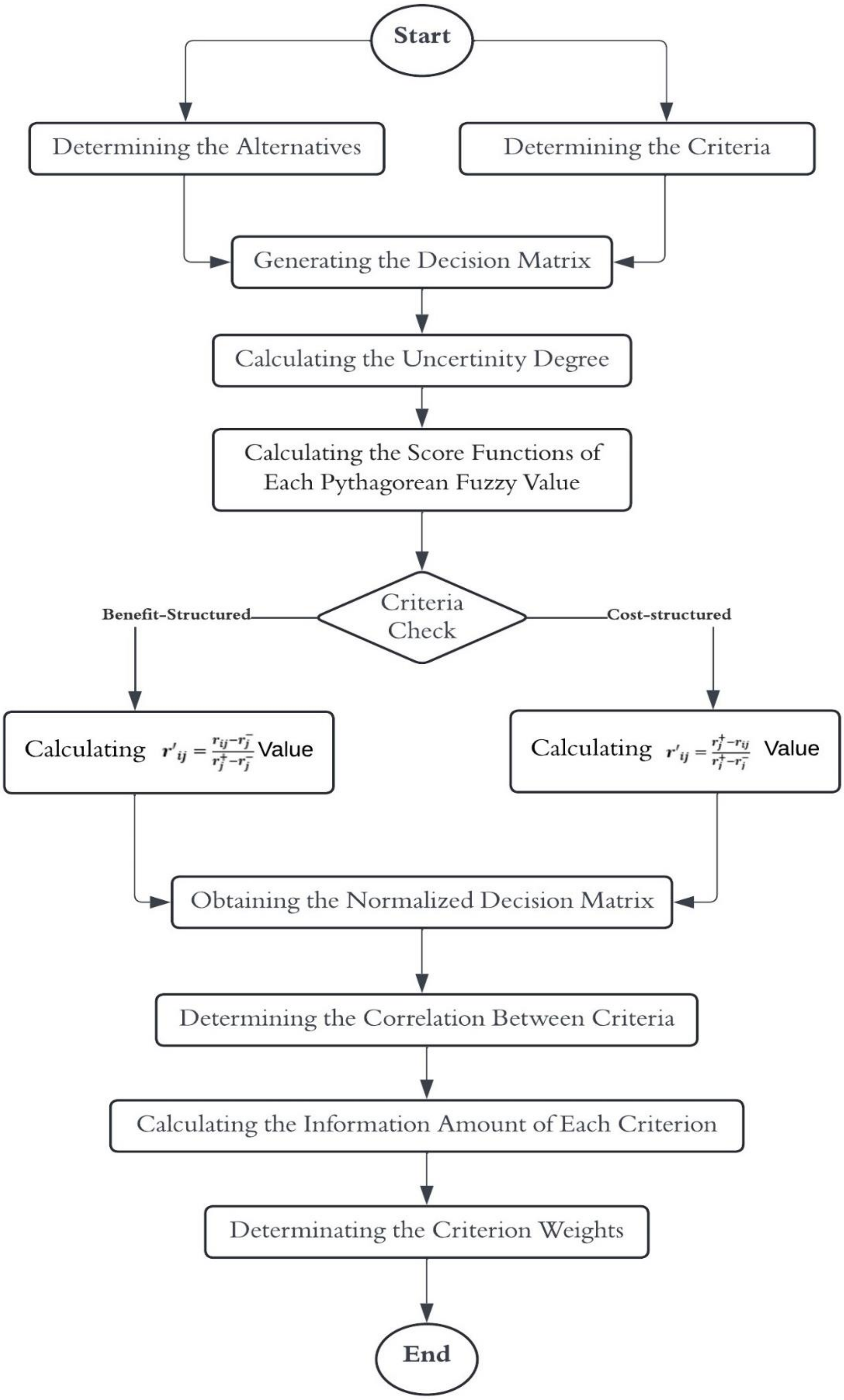


Figure 2. The Process Flow Diagram of the PF-CRITIC Method

The steps of the PF-CRITIC method are presented below [18]:

Step 1: Formation of the Decision Matrix: We construct the initial decision matrix according to Equation (5), where m denotes the number of candidate alternatives and n represents the number of evaluation criteria. Here, for i∈{1, 2,…,m} and j∈{1,2,…,n}), X_ij signifies the performance of the i-th alternative with respect to the j-th criterion.

$$X = \left[x_{ij}\right]_{mxn} = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \tag{5}$$

Step 2: Calculation of Uncertainty Degree: The uncertainty degree of each fuzzy value, denoted as $\mathrm{p_{ij}}(\mu_{ij}, \upsilon_{ij})$, representing the Pythagorean fuzzy value of the i-th alternative with respect to the j-th criterion, is calculated using Equation (6).

$$\prod_{ij} = \sqrt{1 - \mu_{ij}^2 - \upsilon_{ij}^2},\ (i = 1,2,\dots,m; j = 1,2,\dots,n)\ and\ 0 \le \left(\mu_{ij}\right)^2 + \left(\upsilon_{ij}\right)^2 \le 1 \tag{6}$$

Step 3: Calculation of Score Functions for Each Pythagorean Fuzzy Value: For a score matrix $\mathrm{R} = \left(\mathrm{r_{ij}}\right)_{\mathrm{mxn}}$, the score functions for each fuzzy value are calculated as shown in Equation (7).

$$r_{ij} = \mu_{ij}^2 - \upsilon_{ij}^2 - ln\left(1 + \prod_{ij}^2\right),\quad (i = 1,2,\dots,m;\ j = 1,2,\dots,n) \tag{7}$$

Step 4: Normalization Process (Conversion of the Score Matrix R to an Orthonormal Pythagorean Fuzzy Matrix): The transformation process, resulting in the matrix $\mathrm{R}' = \left(\mathrm{r'_{ij}}\right)_{\mathrm{mxn}}$, is conducted using Equation (8) for benefit criteria and Equation (9) for cost criteria.

For benefit criteria; $r'_{ij} = \frac{r_{ij} - r_j^-}{r_j^+ - r_j^-} \quad r_j^- = min_i r_{ij}\ ve\ r_j^+ = max_i r_{ij}$ (8)

For cost criteria; $r'_{ij} = \frac{r_j^+ - r_{ij}}{r_j^+ - r_j^-} \quad r_j^- = min_i r_{ij}\ ve\ r_j^+ = max_i r_{ij}$ (9)

Step 5: Calculation of Criterion Standard Deviations: The standard deviation calculation is determined using Equation (10).

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^{m}\left(r'_{ij} - \bar{r}_j\right)^2}{m}} \qquad \text{Here } \bar{r}_j = \frac{\sum_{i=1}^{m} r'_{ij}}{m} \tag{10}$$

Step 6: Determination of Inter-Criteria Correlation: The correlation value between the j-th criterion and the k-th criterion is calculated using Equation (11).

$$p_{jk} = \frac{\sum_{i=1}^{m}\left(r'_{ij} - \bar{r}_j\right)\left(r'_{ik} - \bar{r}_k\right)}{\sqrt{\sum_{i=1}^{m}\left(r'_{ij} - \bar{r}_j\right)^2 \sum_{i=1}^{m}\left(r'_{ik} - \bar{r}_k\right)^2}} \quad (k, j = 1,2,\dots,n) \tag{11}$$

Step 7: Calculation of Information Amount for Each Criterion: The calculation of the information amount is performed using Equation (12).

$$C_j = \sigma_j \sum_{k=1}^{n}\left(1 - p_{jk}\right) \quad (k, j = 1,2,\dots,n) \tag{12}$$

The larger the value of C_j in Equation (12), the more information a specific criterion contains. Therefore, the weight of this evaluation criterion is greater than the weights of other criteria.

Step 8: Determination of Criterion Weights: Criterion weights are determined using Equation (13).

$$w_j = \frac{C_j}{\sum_{j=1}^{n} C_j} \quad (j = 1,2,\dots,n) \tag{13}$$

4.2. Interval-Valued Pythagorean Fuzzy WASPAS (IVPF-WASPAS) Method

Turskis, Zavadskas, and their colleagues integrated fuzzy logic with the WASPAS method for construction site selection, introducing the fuzzy WASPAS method to the literature in 2015 [19].

The Interval-Valued Pythagorean Fuzzy WASPAS (IVPF-WASPAS) method, introduced to the literature by Ilbahar and Kahraman in 2018, resulted from the adaptation of Pythagorean fuzzy numbers to the classical WASPAS method [20]. They [20] evaluated the performance of retail stores in their study. The process flow diagram of the IVPF-WASPAS method is modeled in Figure 3.

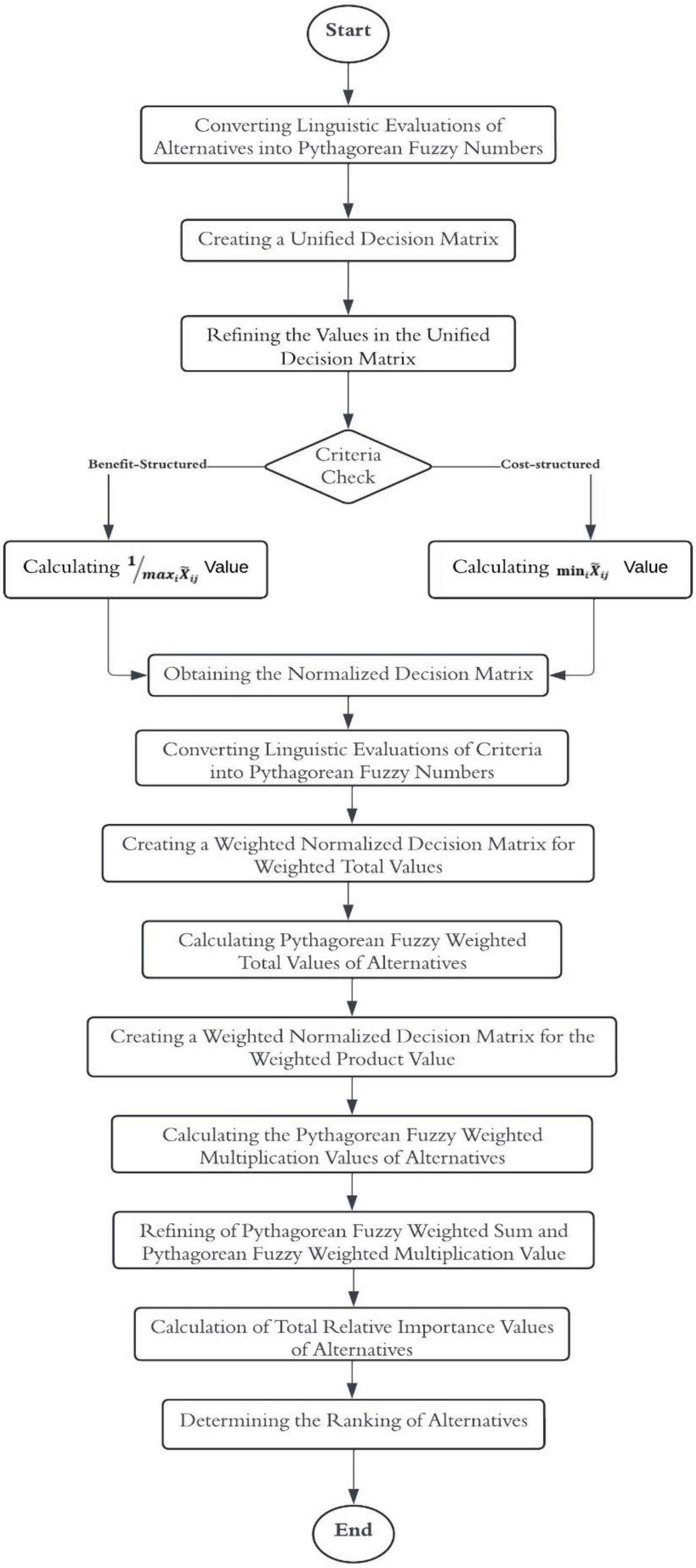


Figure 3. Process Flow Diagram of the IVPF-WASPAS Method

In the IVPF-WASPAS method, the implementation steps are as follows [18]:

Step 1: Formation of the Combined Decision Matrix: Decision-makers gather opinions about alternatives using linguistic expressions. These expressions are converted into Pythagorean fuzzy numbers. The arithmetic mean (IVPFWA) of matrices composed of Pythagorean fuzzy numbers from each expert is calculated using Equation (14), resulting in the creation of the combined decision matrix $\tilde{X}_{ij}$. Here,w_i represents the weight of the criterion.

$$IVPFWA(\tilde{p}_1, \tilde{p}_2, \dots, \tilde{p}_n) = \left(\left[\sum_{i=1}^{n} w_i\mu_i^L, \sum_{i=1}^{n} w_i\mu_i^U\right], \left[\sum_{i=1}^{n} w_i\upsilon_i^L, \sum_{i=1}^{n} w_i\upsilon_i^U\right]\right) \quad (14)$$

Step 2: Obtaining the Normalized Decision Matrix in the Form of Pythagorean Fuzzy Numbers: The defuzzification formula in Equation (15) defuzzifies the values in the resulting combined decision matrix. The "p" value in Equation (15) is an intermediate variable that clarifies the Pythagorean fuzzy number. This formula makes the calculations necessary for the defuzzification process and reduces the uncertainties in the decision matrix. If the criterion is benefit-based after the defuzzification process, Equation (16); If it is cost based, Equation (17) is used. Equation (18) is $1/max_i p_{ij}$; for benefit-based criteria; for cost-based criteria, it is applied for all values in the combined decision matrix using $min_i p_{ij}$ .Thus, the normalized decision matrix ($\bar{\bar{X}}_{ij}$) is obtained.

$$p = \frac{\mu^L + \mu^U + \sqrt{1-(\upsilon^L)^2} + \sqrt{1-(\upsilon^U)^2} + \mu^L\mu^U - \sqrt{\sqrt{1-(\upsilon^L)^2}\sqrt{1-(\upsilon^U)^2}}}{4} \quad (15)$$

$$\text{For benefit criteria; } \bar{\bar{X}}_{ij} = \frac{\tilde{X}_{ij}}{max_i p_{ij}} \quad (16)$$

$$\text{For cost criteria; } \bar{\bar{X}}_{ij} = \frac{\min_i . p_{ij}}{\tilde{X}_{ij}} \quad (17)$$

$$\lambda\tilde{p} = \left(\left[\sqrt{1-(1-(\mu^L)^2)^\lambda}, \sqrt{1-(1-(\mu^U)^2)^\lambda}\right], \left[(\upsilon^L)^\lambda, (\upsilon^U)^\lambda\right]\right) \quad (18)$$

Equation (18) facilitates the transformation of Pythagorean fuzzy numbers using a specific λ coefficient. The λ coefficient is a parameter used during the defuzzification of Pythagorean fuzzy numbers. This transformation reduces uncertainty among the numbers and provides the necessary adjustment for normalization. Equations (16) and (17) then utilize these transformed values to normalize the decision matrix according to benefit and cost criteria. This process aims to make Pythagorean fuzzy numbers comparable and consistent within the decision-making framework.

In summary, the values of the combined decision matrix are defuzzified using Equation 15. Subsequently, Equation 16 is applied for benefit-based criteria using $1/max_i p_{ij}$, and Equation 17 is used for cost-based criteria employing $min_i p_{ij}$. During this process, the p_{ij} values represent previously defuzzified values. The λ values specified in Equation 18 are based on $1/max_i p_{ij}$ and $min_i p_{ij}$. μ^L, μ^U, υ^L, and υ^U denote the Pythagorean fuzzy number values in the combined decision matrix. These computational steps ensure the comparability and normalization of values in decision-making processes involving Pythagorean fuzzy numbers.

Step 3: Conversion of Linguistic Evaluations for Criteria into Pythagorean Fuzzy Numbers: The linguistic expressions regarding the importance levels of criteria provided by decision-makers are transformed into Pythagorean fuzzy numbers.

Step 4: Obtaining the Pythagorean Fuzzy Weighted Sum Values of Alternatives: The weighted sum matrix, which is the first part of the WASPAS method, is obtained through Equation (19). Here, w_j represents the weight of the criterion.

$$\tilde{Q}_i^{(1)} = \sum_{j=1}^{n} \bar{\bar{X}}_{ij} \,.\, w_j \tag{19}$$

Before Equation (19) can be applied, the values in the normalized decision matrix must first be multiplied by the criterion weights through Equation (20). These values are then summed with each other in Equation (21) to obtain the Pythagorean fuzzy weighted sum values of alternatives.

$$\tilde{p}_1 \otimes \tilde{p}_2 = \left([\mu_1^L \mu_2^L, \mu_1^U \mu_2^U], \left[\sqrt{(\upsilon_1^L)^2 + (\upsilon_2^L)^2 - (\upsilon_1^L)^2(\upsilon_2^L)^2}, \sqrt{(\upsilon_1^U)^2 + (\upsilon_2^U)^2 - (\upsilon_1^U)^2(\upsilon_2^U)^2}\right]\right) \tag{20}$$

$$\tilde{p}_1 \oplus \tilde{p}_1 = \left(\left[\sqrt{(\mu_1^L)^2 + (\mu_2^L)^2 - (\mu_1^L)^2(\mu_2^L)^2}, \sqrt{(\mu_1^U)^2 + (\mu_2^U)^2 - (\mu_1^U)^2(\mu_2^U)^2}\right], [\upsilon_1^L \upsilon_2^L, \upsilon_1^U \upsilon_2^U]\right) \tag{21}$$

Step 5: Obtaining the Pythagorean Fuzzy Weighted Product Values of Alternatives: The weighted product matrix of the WASPAS method is obtained using Equation (22). Before Equation (22) can be applied, Equation (23) is first applied to the values in the normalized decision matrix and the criterion weights. Then, by multiplying these values with each other in Equation (20), the Pythagorean fuzzy weighted product values of alternatives are obtained.

$$\tilde{Q}_i^{(2)} = \prod_{j=1}^{n} \left(\bar{\bar{X}}_{ij}\right)^{w_j} \tag{22}$$

$$p^{\lambda} = \left(\left[(\mu^L)^{\lambda}, (\mu^U)^{\lambda}\right], \left[\sqrt{1 - (1 - (\upsilon^L)^2)^{\lambda}}, \sqrt{1 - (1 - (\upsilon^U)^2)^{\lambda}}\right]\right) \tag{23}$$

Step 6: Determination of the Total Relative Importance Values of Alternatives: The Pythagorean fuzzy weighted sum and weighted product values are normalized using Equation (15). According to the WASPAS method, the weighted sum values and weighted product values of alternatives are integrated through Equation (24). Thus, the total relative importance value (Q_i) of alternatives is obtained, providing a single value for decision-making. Subsequently, the obtained values are weighted and summed using the λ coefficient. The λ coefficient represents the importance levels assigned to two values and should take a value between 0 and 1.

$$Q_i = \lambda Q_i^{(1)} + (1 - \lambda) Q_i^{(2)} = \lambda \sum_{j=1}^{n} \bar{X}_{ij} . w_j + (1 - \lambda) \prod_{j=1}^{n} (\bar{X}_{ij})^{w_j} \tag{24}$$

Step 7: Determination of Alternative Rankings: Alternatives are ranked based on their total relative importance values. The alternative with the highest total relative importance value is preferred.

5. Case Study

The study employed Fuzzy Multi-Criteria Decision Making (FMCDM) methods, specifically utilizing PF-CRITIC for criteria weighting and IVPF-WASPAS for simultaneous alternative ranking. Three decision-makers assessed the following alternatives: "Home Healthcare Services", "Medical Centers", "Psychological Counseling Centers", "Elderly Services", "Healthy Nutrition Support" and "Mobile Healthcare Services".

Five criteria were established to evaluate the digitization of healthcare services: "Urgency and Importance Level", "Social Needs and Demands", "Accessibility and Inclusivity", "Efficiency and Cost Effectiveness" and "Technological Infrastructure and Capabilities".

Some criteria considered in the decision-making process are benefit-oriented depending on the problem's nature, while others may focus on cost. Decision-makers aim to maximize benefit-oriented criteria and minimize cost-oriented ones.

The reason these criteria—"Urgency and Importance Level," "Social Needs and Demands", "Accessibility and Inclusivity", "Efficiency and Cost Effectiveness" and "Technological Infrastructure and Capabilities"—are benefit-oriented is due to the high demand from decision-makers in achieving their objectives. Essentially, decision-makers perceive these criteria as representing positive attributes and seek to maximize their value.

5.1 The Implementation of the PF-CRITIC Method

Step 1: Table 1 displays the Pythagorean Fuzzy values utilized for weighting the criteria in the PF-CRITIC method.

Table 1. The Nine-Point Pythagorean Fuzzy Linguistic Variables Scale Used to Evaluate Alternatives in Terms of Criteria [23]

Language Terms	The Corresponding Pythagorean Fuzzy Number $(\boldsymbol{u}.\boldsymbol{v})$
Extremely Low (EL)	(0.10,0.99)
Very Little (VL)	(0.10,0.97)
Little (L)	(0.25,0.92)
Middle Little (ML)	(0.40,0.87)
Middle (M)	(0.50,0.80)
Middle High (MH)	(0.60,0.71)
Big (B)	(0.70,0.60)
Very Tall (VT)	(0.80,0.44)
Tremendously High (TH)	(0.10,0.00)

Utilizing Equation (5), Tables 2, 3, and 4 present the fuzzy decision matrices created for each decision-maker.

Table 2. Fuzzy Decision Matrix for Decision Maker-1

Alternatives	Criteria				
	Urgency and Importance Level	Social Needs and Demands	Accessibility and Inclusivity	Efficiency and Cost Effectiveness	Technological Infrastructure and Capabilities
Home Health Services	M	B	B	B	MH
Medical Centers	B	TH	VT	B	VT
Psychological Counseling Centers	MH	B	VT	VT	MH
Elderly Services	MH	B	B	B	M
Healthy Nutrition Support	M	M	B	VT	B
Mobile Health Services	VT	VT	TH	TH	TH

Table 3. Fuzzy Decision Matrix for Decision Maker-2

Alternatives	Criteria				
	Urgency and Importance Level	Social Needs and Demands	Accessibility and Inclusivity	Efficiency and Cost Effectiveness	Technological Infrastructure and Capabilities
Home Health Services	L	B	VL	M	VL
Medical Centers	M	ML	VT	ML	L
Psychological Counseling Centers	M	B	VL	VT	L
Elderly Services	M	VT	L	MH	M
Healthy Nutrition Support	MH	B	MH	B	ML
Mobile Health Services	B	MH	EL	VT	TH

Table 4. Fuzzy Decision Matrix for Decision Maker-3

Alternatives	Criteria				
	Urgency and Importance Level	Social Needs and Demands	Accessibility and Inclusivity	Efficiency and Cost Effectiveness	Technological Infrastructure and Capabilities
Home Health Services	EL	ML	M	VL	TH
Medical Centers	EL	VT	MH	ML	M
Psychological Counseling Centers	EL	VL	B	EL	ML
Elderly Services	EL	B	VT	VT	MH
Healthy Nutrition Support	EL	MH	EL	B	VT
Mobile Health Services	VL	M	ML	MH	B

In Table 5, the Pythagorean fuzzy number versions of the linguistic terms used in the fuzzy decision matrix for decision maker-1 in Table 2 have been presented. Similar procedures were applied for decision maker-2 and decision maker-3 using Tables 3 and 4 respectively.

Table 5. The Pythagorean Fuzzy Number Counterparts of The Fuzzy Decision Matrix for Decision Maker-1

Alternatives	Criteria									
	Urgency and Importance Level		Social Needs and Demands		Accessibility and Inclusivity		Efficiency and Cost Effectiveness		Technological Infrastructure and Capabilities	
Home Health Services	0.5	0.8	0.7	0.6	0.7	0.6	0.7	0.6	0.6	0.71
Medical Centers	0.7	0.6	0.1	0	0.8	0.44	0.7	0.6	0.8	0.44
Psychological Counseling Centers	0.6	0.71	0.7	0.6	0.8	0.44	0.8	0.44	0.6	0.71
Elderly Services	0.6	0.71	0.7	0.6	0.7	0.6	0.7	0.6	0.5	0.8
Healthy Nutrition Support	0.5	0.8	0.5	0.8	0.7	0.6	0.8	0.44	0.7	0.6
Mobile Health Services	0.8	0.44	0.8	0.44	0.1	0	0.1	0	0.1	0

Step 2: The uncertainty degree of each Pythagorean fuzzy value has been calculated using Equation (6). The uncertainty matrix calculated for decision maker-1 is presented in Table 6. Similar procedures have been applied for decision makers 2 and 3 as well.

Table 6. Uncertainty Matrix for Decision Maker-1

Alternatives	Criteria				
	Urgency and Importance Level	Social Needs and Demands	Accessibility and Inclusivity	Efficiency and Cost Effectiveness	Technological Infrastructure and Capabilities
Home Health Services	0.332	0.387	0.387	0.387	0.369
Medical Centers	0.387	0.995	0.408	0.387	0.408
Psychological Counseling Centers	0.369	0.387	0.408	0.408	0.369
Elderly Services	0.369	0.387	0.387	0.387	0.332
Healthy Nutrition Support	0.332	0.332	0.387	0.408	0.387
Mobile Health Services	0.408	0.408	0.995	0.995	0.995

Step 3: The score function of each Pythagorean fuzzy value has been found using Equation (7). The score matrix for decision maker-1 is presented in Table 7. Similar procedures have been applied for decision makers 2 and 3 as well.

Table 7. Score Matrix for Decision Maker-1

Alternatives	**Criteria**				
	Urgency and Importance Level	Social Needs and Demands	Accessibility and Inclusivity	Efficiency and Cost Effectiveness	Technological Infrastructure and Capabilities
Home Health Services	-0.494	-0.010	-0.010	-0.010	-0.272
Medical Centers	-0.010	-0.678	0.292	-0.010	0.292
Psychological Counseling Centers	-0.272	-0.010	0.292	0.292	-0.272
Elderly Services	-0.272	-0.010	-0.010	-0.010	-0.494
Healthy Nutrition Support	-0.494	-0.494	-0.010	0.292	-0.010
Mobile Health Services	0.292	0.292	-0.678	-0.678	-0.678
Maximum	0.292	0.292	0.292	0.292	0.292
Minimum	-0.494	-0.678	-0.678	-0.678	-0.678

Step 4: The score matrix has been transformed into an orthonormal Pythagorean fuzzy matrix for the benefit-oriented criteria using Equation (8). The orthonormal Pythagorean fuzzy matrix (normalization matrix) for decision maker-1 is presented in Table 8. Similar procedures have been applied for decision makers 2 and 3 as well.

Table 8. Orthonormal Pythagorean Fuzzy Matrix for Decision Maker-1 (Normalization matrix)

	Criteria				
Alternatives	Urgency and Importance Level	Social Needs and Demands	Accessibility and Inclusivity	Efficiency and Cost Effectiveness	Technological Infrastructure and Capabilities
Home Health Services	0.000	0.689	0.689	0.689	0.419
Medical Centers	0.616	0.000	1.000	0.689	1.000
Psychological Counseling Centers	0.283	0.689	1.000	1.000	0.419
Elderly Services	0.283	0.689	0.689	0.689	0.189
Healthy Nutrition Support	0.000	0.189	0.689	1.000	0.689
Mobile Health Services	1.000	1.000	0.000	0.000	0.000

Step 5: According to the values in Table 8, the standard deviations of the criteria for decision maker-1 are determined using Equation (10). In Table 9, the standard deviation values for decision maker-1 were calculated using the "STDEV ()" function in Excel. Similar procedures have been applied for decision makers 2 and 3 as well.

Table 9. The Standard Deviation Values of The Criteria for Decision Maker-1

Criteria	Urgency and Importance Level	Social Needs and Demands	Accessibility and Inclusivity	Efficiency and Cost Effectiveness	Technological Infrastructure and Capabilities
σ	0.353	0.340	0.334	0.334	0.324

Step 6: The correlation value between criteria is calculated using Equation (11). To apply Equation (11), the correlation matrix for decision maker-1 was created using the "CORREL ()" function in Excel, and it is presented in Table 10. Similar procedures have been applied for decision makers 2 and 3 as well.

Table 10. Correlation Matrix for Decision Maker-1

Criteria	Urgency and Importance Level	Social Needs and Demands	Accessibility and Inclusivity	Efficiency and Cost Effectiveness	Technological Infrastructure and Capabilities
Urgency and Importance Level	1.000	0.287	-0.545	-0.817	-0.291
ocial Needs and Demands	0.287	1.000	-0.645	-0.558	-0.961
Accessibility and Inclusivity	-0.545	-0.645	1.000	0.855	0.726
Efficiency and Cost Effectiveness	-0.817	-0.558	0.855	1.000	0.577
Technological Infrastructure and Capabilities	-0.291	-0.961	0.726	0.577	1.000

Step 7: The amount of information provided by each criterion (useful information value) is calculated using Equation (12). The information amount of the criteria for decision maker-1 is presented in Table 11. Similar procedures have been applied for decision makers 2 and 3 as well.

Table 11. Information Value of Criteria for Decision Maker-1

Criteria	Urgency and Importance Level	Social Needs and Demands	Accessibility and Inclusivity	Efficiency and Cost Effectiveness	Technological Infrastructure and Capabilities
c	1.892	1.997	1.204	1.315	1.281

Step 8: The weights of the criteria for each decision-maker are calculated using Equation (13). The weights of the criteria for decision maker-1 have been presented in Table 12 using Equation (13). Similar procedures have been applied for decision makers 2 and 3 as well.

Table 12. Weights of Criteria for Decision Maker-1

Criteria	Urgency and Importance Level	Social Needs and Demands	Accessibility and Inclusivity	Efficiency and Cost Effectiveness	Technological Infrastructure and Capabilities
w	0.246	0.260	0.157	0.171	0.167
Prioritization	2	1	5	3	4

Upon examining the results obtained from the PF-CRITIC method:

For decision maker-1, the importance ranking of criteria is as follows: "Social Needs and Demands" > "Urgency and Importance Level" > "Efficiency and Cost Effectiveness" > "Technological Infrastructure and Capabilities" > "Accessibility and Inclusivity".

For decision maker-2, the importance ranking of criteria is as follows: "Accessibility and Inclusivity" > "Social Needs and Demands" > "Efficiency and Cost Effectiveness" > "Technological Infrastructure and Capabilities" > "Urgency and Importance Level".

For decision maker-3, the importance ranking of criteria is as follows: "Accessibility and Inclusivity" > "Urgency and Importance Level" > "Technological Infrastructure and Capabilities" > "Social Needs and Demands" > "Efficiency and Cost Effectiveness".

As observed, for decision maker-2 and decision maker-3, the "Accessibility and Inclusivity" criterion is the most important factor, while for decision maker-1, this criterion is determined as the least prioritized factor. For decision maker-1 and decision maker-3, the second priority factor is "Urgency and Importance Level", whereas for decision maker-2, this criterion is determined as the least prioritized factor. For decision maker-1 and decision maker-2, the third priority factor is "Efficiency and Cost Effectiveness", while for decision maker-3, this criterion is determined as the least prioritized factor. For decision maker-1 and decision maker-2, the fourth priority factor is "Technological Infrastructure and Capabilities".

5.2 The Implementation of the IVPF-WASPAS Method

Step 1: The comparison scale used in linguistic evaluations about alternatives in the IVPF-WASPAS method is provided in Table 13.

Table 13. Comparison Scale for Evaluating Alternatives [17]

Linguistic Terms		**IVPF Numbers**			
		μ^L	μ^U	v^L	v^U
CCI	Extremely good	0.8	0.9	0.1	0.2
CI	Very good	0.7	0.8	0.2	0.3
I	Good	0.6	0.7	0.3	0.4
O	Fair	0.5	0.6	0.3	0.5
K	Poor	0.3	0.4	0.6	0.7
CK	Very poor	0.2	0.3	0.7	0.8
CCK	Extremely poor	0.1	0.2	0.8	0.9

Decision maker-1, decision maker-2, and decision maker-3 provided their opinions regarding the alternatives using linguistic expressions. Tables 14, 15, and 16 present the relevant information for each decision maker, respectively.

Table 14. Linguistic Terms Used by Decision Maker-1 to Rank The Importance of Alternatives

Alternatives	**Criteria**				
	Urgency and Importance Level	Social Needs and Demands	Accessibility and Inclusivity	Efficiency and Cost Effectiveness	Technological Infrastructure and Capabilities
Home Health Services	I	CI	CCI	CI	CCI
Medical Centers	CI	CCI	CI	CI	CCI
Psychological Counseling Centers	I	I	CI	O	I
Elderly Services	CI	CI	I	O	CI
Healthy Nutrition Support	I	O	CI	O	I
Mobile Health Services	CCI	CI	CCI	CCI	CCI

Table 15. Linguistic Terms Used by Decision Maker-2 to Rank the Importance of Alternatives

Alternatives	**Criteria**				
	Urgency and Importance Level	Social Needs and Demands	Accessibility and Inclusivity	Efficiency and Cost Effectiveness	Technological Infrastructure and Capabilities
Home Health Services	CI	I	O	K	CK
Medical Centers	CCI	CCI	CI	CCK	O
Psychological Counseling Centers	O	I	CI	I	CK
Elderly Services	CI	CI	CI	O	CK
Healthy Nutrition Support	O	O	O	CI	I
Mobile Health Services	CCI	CI	CCI	CCI	CCI

Table 16. Linguistic Terms Used by Decision Maker-3 to Rank the Importance of Alternatives

Alternatives	**Criteria**				
	Urgency and Importance Level	Social Needs and Demands	Accessibility and Inclusivity	Efficiency and Cost Effectiveness	Technological Infrastructure and Capabilities
Home Health Services	CCK	CI	CK	CCI	CI
Medical Centers	CK	CCI	CCI	CI	CK
Psychological Counseling Centers	CI	K	CK	CK	CCI
Elderly Services	CI	CK	CCI	CCI	CI
Healthy Nutrition Support	CK	CI	CI	CK	CCK
Mobile Health Services	CK	I	CK	CI	K

The linguistic expressions in Tables 14, 15 and 16 should be converted into Pythagorean fuzzy numbers. The linguistic expressions' Pythagorean fuzzy counterparts for decision maker-1 are presented in Table 17. Table 18 presents the combined decision matrix values $\tilde{X}_{ij}$ for decision makers using Equation (14).

Table 17. Pythagorean Fuzzy Number Equivalents of Linguistic Expressions for Decision Maker-1

Criteria	Urgency and Importance Level				Social Needs and Demands				Accessibility and Inclusivity				Efficiency and Cost Effectiveness				Technological Infrastructure and Capabilities			
Alternatives	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U
Home Health Services	0.6	0.7	0.3	0.4	0.7	0.8	0.2	0.3	0.8	0.9	0.1	0.2	0.7	0.8	0.2	0.3	0.8	0.9	0.1	0.2
Medical Centers	0.7	0.8	0.2	0.3	0.8	0.9	0.1	0.2	0.7	0.8	0.2	0.3	0.7	0.8	0.2	0.3	0.8	0.9	0.1	0.2
Psychological Counseling Centers	0.6	0.7	0.3	0.4	0.6	0.7	0.3	0.4	0.7	0.8	0.2	0.3	0.5	0.6	0.3	0.5	0.6	0.7	0.3	0.4
Elderly Services	0.7	0.8	0.2	0.3	0.7	0.8	0.2	0.3	0.6	0.7	0.3	0.4	0.5	0.6	0.3	0.5	0.7	0.8	0.2	0.3
Healthy Nutrition Support	0.6	0.7	0.3	0.4	0.5	0.6	0.3	0.5	0.7	0.8	0.2	0.3	0.5	0.6	0.3	0.5	0.6	0.7	0.3	0.4
Mobile Health Services	0.8	0.9	0.1	0.2	0.7	0.8	0.2	0.3	0.8	0.9	0.1	0.2	0.8	0.9	0.1	0.2	0.8	0.9	0.1	0.2

Table 18. Combined Decision Matrix

Criteria	Urgency and Importance Level				Social Needs and Demands				Accessibility and Inclusivity				Efficiency and Cost Effectiveness				Technological Infrastructure and Capabilities			
Alternatives	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U
Home Health Services	0.269	0.331	0.290	0.352	0.422	0.485	0.146	0.209	0.333	0.406	0.288	0.392	0.302	0.355	0.169	0.221	0.293	0.343	0.157	0.207
Medical Centers	0.331	0.393	0.228	0.290	0.505	0.568	0.063	0.126	0.533	0.606	0.120	0.192	0.247	0.300	0.224	0.276	0.244	0.294	0.191	0.256
Psychological Counseling Centers	0.382	0.444	0.163	0.239	0.327	0.390	0.241	0.304	0.381	0.453	0.272	0.345	0.236	0.288	0.219	0.288	0.278	0.328	0.172	0.222
Elderly Services	0.434	0.496	0.124	0.186	0.356	0.419	0.212	0.275	0.517	0.590	0.135	0.208	0.308	0.360	0.126	0.216	0.276	0.326	0.173	0.223
Healthy Nutrition Support	0.264	0.326	0.280	0.356	0.350	0.413	0.172	0.281	0.445	0.517	0.177	0.281	0.255	0.308	0.199	0.268	0.207	0.257	0.243	0.293
Mobile Health Services	0.355	0.417	0.203	0.265	0.424	0.488	0.143	0.206	0.428	0.501	0.225	0.297	0.404	0.456	0.068	0.120	0.307	0.357	0.143	0.193

Step 2: The formula that provides Equation (15) defused the values in the integrated decision matrix. After the defuzzification process, Equation (16) was used since all criteria are utility-based. The defuzzified values obtained are presented in Table 19. Equation (18) was applied to all values $1/max_i\tilde{X}_{ij}$ in the combined decision matrix for benefit-based criteria. Thus, we obtained the normalized decision matrix $(\bar{\bar{X}}_{ij})$ for decision makers, which is presented in Table 20.

Table 19. Defuzzified Values

Alternatives	Urgency and Importance Level	Social Needs and Demands	Accessibility and Inclusivity	Efficiency and Cost Effectiveness	Technological Infrastructure and Capabilities
Home Health Services	0.409	0.524	0.453	0.436	0.430
Medical Centers	0.455	0.589	0.612	0.397	0.396
Psychological Counseling Centers	0.493	0.452	0.489	0.390	0.419
Elderly Services	0.533	0.473	0.599	0.441	0.418
Healthy Nutrition Support	0.406	0.470	0.541	0.403	0.370
Mobile Health Services	0.473	0.526	0.527	0.510	0.440
1/Maximum	1.875	1.699	1.633	1.962	2.275

Table 20. Normalized Decision Matrix in Pythagorean Fuzzy Numbers

Criteria	Urgency and Importance Level				Social Needs and Demands				Accessibility and Inclusivity				Efficiency and Cost Effectiveness				Technological Infrastructure and Capabilities			
Alternatives	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U
Home Health Services	0.362	0.442	0.098	0.141	0.532	0.605	0.038	0.070	0.419	0.505	0.131	0.216	0.414	0.482	0.031	0.052	0.430	0.498	0.015	0.028
Medical Centers	0.442	0.519	0.062	0.098	0.627	0.696	0.009	0.030	0.649	0.725	0.031	0.068	0.341	0.411	0.053	0.080	0.361	0.431	0.023	0.045
Psychological Counseling Centers	0.506	0.581	0.033	0.068	0.418	0.495	0.089	0.132	0.475	0.560	0.119	0.175	0.326	0.395	0.051	0.087	0.409	0.478	0.018	0.032
Elderly Services	0.569	0.641	0.020	0.043	0.453	0.529	0.072	0.111	0.632	0.709	0.038	0.077	0.421	0.489	0.017	0.049	0.407	0.475	0.019	0.033
Healthy Nutrition Support	0.356	0.436	0.092	0.144	0.446	0.522	0.050	0.116	0.550	0.631	0.059	0.125	0.352	0.421	0.042	0.076	0.308	0.379	0.040	0.061
Mobile Health Services	0.473	0.549	0.050	0.083	0.535	0.608	0.037	0.068	0.530	0.613	0.087	0.138	0.543	0.606	0.005	0.016	0.449	0.516	0.012	0.024

Step 3: Utilizing the comparison scale in Table 21, the joint linguistic assessments of decision makers regarding the importance levels of the criteria are presented in Table 22. The combined criterion weights for decision makers, expressed as Pythagorean fuzzy numbers, are provided in Table 23.

Table 21. Linguistic Terms for Rating the Importance of Criteria [21]

Linguistic Terms	$\boldsymbol{\mu^L}$	$\boldsymbol{\mu^U}$	$\boldsymbol{\upsilon^L}$	$\boldsymbol{\upsilon^U}$
Very important (VI)	0.70	0.90	0.06	0.26
Important (I)	0.54	0.74	0.22	0.42
Medium (M)	0.38	0.58	0.38	0.58
Unimportant (U)	0.22	0.42	0.54	0.74
Very unimportant (VU)	0.06	0.26	0.70	0.90

Table 22. Linguistic Terms Used by Decision Makers to Rate the Importance of Criteria

Criteria	**Linguistic Terms**
Urgency and Importance Level	I
Social Needs and Demands	I
Accessibility and Inclusivity	M
Efficiency and Cost Effectiveness	I
Technological Infrastructure and Capabilities	VI

Table 23. Criterion Weights in The Form of Pythagorean Numbers

Criteria				
Urgency and Importance Level	0.54	0.74	0.22	0.42
Social Needs and Demands	0.54	0.74	0.22	0.42
Accessibility and Inclusivity	0.38	0.58	0.38	0.58
Efficiency and Cost Effectiveness	0.54	0.74	0.22	0.42
Technological Infrastructure and Capabilities	0.7	0.9	0.06	0.26

Step 4: The values in the normalized decision matrix have been multiplied by the criteria weights using Equation (20). Table 24 provides the resulting weighted normalized decision matrix for the weighted total value. Subsequently, we obtained the Pythagorean fuzzy weighted total values of the alternatives by summing these values using Equation (21), as shown in Table 25.

Table 24. Weighted Normalized Decision Matrix for Weighted Total Value

Criteria	Urgency and Importance Level				Social Needs and Demands				Accessibility and Inclusivity				Efficiency and Cost Effectiveness				Technological Infrastructure and Capabilities			
Alternatives	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U
Home Health Services	0.195	0.327	0.240	0.439	0.287	0.447	0.223	0.425	0.159	0.293	0.399	0.606	0.224	0.356	0.222	0.423	0.301	0.448	0.062	0.261
Medical Centers	0.239	0.384	0.228	0.429	0.339	0.515	0.220	0.421	0.247	0.421	0.381	0.583	0.184	0.304	0.226	0.426	0.253	0.388	0.064	0.264
Psychological Counseling Centers	0.273	0.430	0.222	0.425	0.226	0.366	0.236	0.437	0.181	0.325	0.396	0.597	0.176	0.292	0.225	0.427	0.286	0.430	0.063	0.262
Elderly Services	0.307	0.475	0.221	0.422	0.245	0.391	0.231	0.432	0.240	0.411	0.382	0.583	0.228	0.362	0.221	0.422	0.285	0.428	0.063	0.262
Healthy Nutrition Support	0.192	0.323	0.238	0.440	0.241	0.386	0.225	0.433	0.209	0.366	0.384	0.589	0.190	0.312	0.224	0.426	0.215	0.341	0.072	0.267
Mobile Health Services	0.255	0.406	0.225	0.427	0.289	0.450	0.223	0.425	0.202	0.355	0.388	0.591	0.293	0.448	0.220	0.420	0.314	0.464	0.061	0.261

Table 25. Pythagorean Fuzzy Weighted Total Values of The Alternatives

Alternatives	μ^L	μ^U	υ^L	υ^U
Home Health Services	0.535	0.849	0.000	0.012
Medical Centers	0.575	0.912	0.000	0.012
Psychological Counseling Centers	0.521	0.833	0.000	0.012
Elderly Services	0.587	0.928	0.000	0.012
Healthy Nutrition Support	0.470	0.775	0.000	0.013
Mobile Health Services	0.611	0.954	0.000	0.012

Step 5: Equation (23) calculated the weighted normalized decision matrix using the values from the normalized decision matrix and the criterion weights. Table 26 displays the resulting weighted normalized decision matrix for the weighted product value.

Subsequently, Table 27 presents the Pythagorean fuzzy weighted product values of the alternatives obtained by multiplying these values together through Equation (20).

Table 26. Weighted Normalized Decision Matrix for The Weighted Product Value

Criteria	Urgency and Importance Level				Social Needs and Demands				Accessibility and Inclusivity				Efficiency and Cost Effectiveness				Technological Infrastructure and Capabilities			
Alternatives	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U	μ^L	μ^U	υ^L	υ^U
Home Health Services	0.578	0.546	0.046	0.092	0.711	0.689	0.018	0.045	0.718	0.673	0.081	0.166	0.621	0.582	0.014	0.034	0.554	0.534	0.004	0.014
Medical Centers	0.643	0.616	0.029	0.064	0.777	0.764	0.004	0.019	0.848	0.830	0.019	0.052	0.560	0.518	0.025	0.052	0.490	0.469	0.006	0.023
Psychological Counseling Centers	0.692	0.669	0.016	0.044	0.625	0.594	0.042	0.086	0.754	0.714	0.074	0.134	0.546	0.503	0.024	0.056	0.535	0.515	0.004	0.017
Elderly Services	0.738	0.720	0.009	0.028	0.652	0.624	0.034	0.072	0.840	0.819	0.024	0.059	0.627	0.589	0.008	0.032	0.533	0.512	0.005	0.017
Healthy Nutrition Support	0.573	0.541	0.043	0.094	0.647	0.618	0.024	0.075	0.797	0.766	0.036	0.096	0.569	0.527	0.020	0.049	0.438	0.418	0.010	0.031
Mobile Health Services	0.667	0.642	0.024	0.054	0.714	0.692	0.017	0.044	0.786	0.753	0.054	0.105	0.719	0.690	0.002	0.010	0.571	0.551	0.003	0.012

Table 27. Pythagorean Fuzzy Weighted Product Values of Alternatives

Alternatives	μ^L	μ^U	υ^L	υ^U
Home Health Services	0.101	0.079	0.096	0.198
Medical Centers	0.116	0.095	0.044	0.101
Psychological Counseling Centers	0.095	0.073	0.089	0.175
Elderly Services	0.135	0.111	0.043	0.104
Healthy Nutrition Support	0.074	0.056	0.065	0.164
Mobile Health Services	0.154	0.127	0.061	0.127

Step 6: Equation (15) defuzzifies the Pythagorean fuzzy weighted sum and the Pythagorean fuzzy weighted product. Table 28 presents the defuzzified weighted sum values, and Table 29 shows the defuzzified weighted product values. According to the WASPAS method, Equation (24) integrates the weighted sum values of the alternatives with the weighted product values, and Table 30 presents the total relative importance values of the alternatives. A λ coefficient value of 0.5 is assigned at this stage.

Table 28. Defuzzified Weighted Sum Values

Alternatives	**Weighted Sum Value**
Home Health Services	0.710
Medical Centers	0.753
Psychological Counseling Centers	0.697
Elderly Services	0.765
Healthy Nutrition Support	0.653
Mobile Health Services	0.787

Table 29. Defuzzified Weighted Product Values

Alternatives	Weighted Product Value
Home Health Services	0.294
Medical Centers	0.305
Psychological Counseling Centers	0.291
Elderly Services	0.314
Healthy Nutrition Support	0.282
Mobile Health Services	0.324

Table 30. Total Relative Importance Values of Alternatives

Alternatives	Total Relative Importance Value	Ranking
Home Health Services	0.502	4
Medical Centers	0.529	3
Psychological Counseling Centers	0.494	5
Elderly Services	0.540	2
Healthy Nutrition Support	0.467	6
Mobile Health Services	0.556	1

Step 7: The alternatives have been ranked considering their total relative importance values. The alternative with the highest total relative importance value is considered the most suitable candidate. The rankings of digitalization alternatives for decision makers are presented in Figure 4.

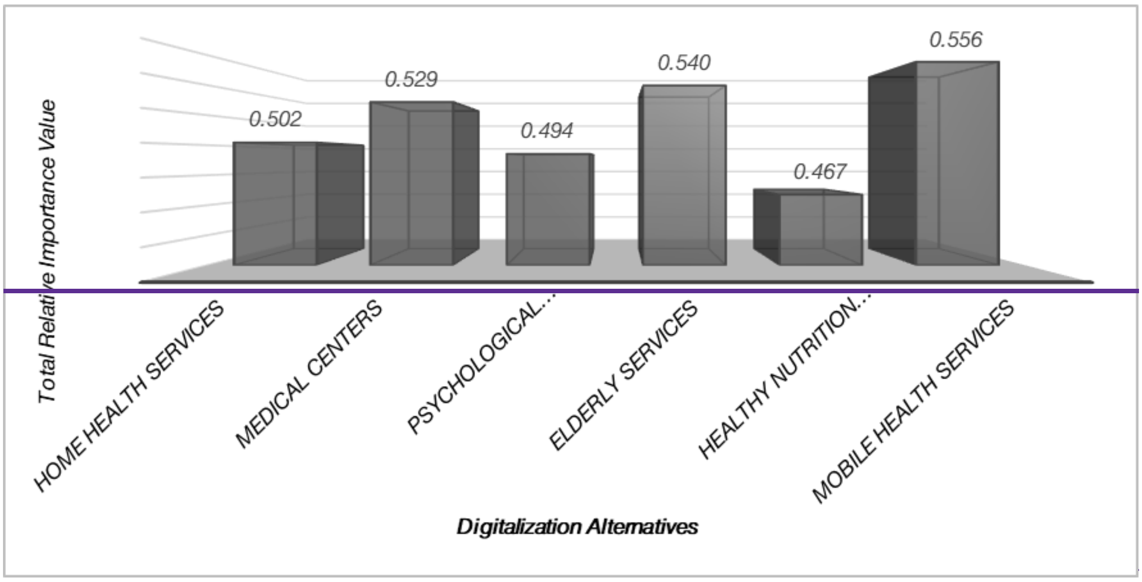


Figure 4. The Rankings of Digitalization Alternatives

Upon examining Figure 4, the importance ranking of alternatives for decision makers is as follows: "Mobile Health Services" > "Elderly Services" > "Medical Centers" > "Home Health Services" > "Psychological Counseling Centers" > "Healthy Nutrition Support".

5.3. The Implementation of the Sensitivity Analysis

To test the validity of the proposed integrated model, a comprehensive sensitivity analysis examined the impact of variations in different criteria weights on the ranking results. This analysis involved creating 50 scenarios to analyze how modifications in criterion weights affected the new ranking of alternatives. Each scenario adjusted the weight of a specific criterion by 10%, while the weights of the remaining criteria were adjusted to maintain a total sum of 1, as recommended by Görçün et al. [22]. The new weight values for each criterion were calculated using Equations (25), (26), and (27) respectively.

$$w_{nv}^{1} = w_{pv}^{1} - \left(w_{pv}^{1}.\varsigma_{v}\right) \tag{25}$$

$$w_{rfv}^{2} = \frac{\left(w_{pv}^{1} - w_{nv}^{1}\right)}{n-1} + w_{pv}^{2} \tag{26}$$

$$w_{nv}^{1} + \sum w_{rfv}^{2} = 1 \tag{27}$$

In Equation (25), w_{pv}^{1}, denotes the original value of the criterion to be reduced in weight; ς_{v} represents the degree of change in percentage terms (10%, 20%...100%); and w_{nv}^{1} signifies the new value of the modified weight of the factor. In Equation (26), w_{pv}^{2} symbolizes the original value of the remaining criterion; n denotes the number of criteria; and w_{rfv}^{2} represents the new value of the remaining criterion. Equation (27) expresses the constraint that the sum of the modified criterion weights must equal 1.

Within the study's scope, we systematically reduced the weights of each factor obtained from the PF-CRITIC method by 10% increments until each factor's weight reached 0, while ensuring the total weight sum of all factors remained at 1. For instance, starting with the "Urgency and Importance Level" criterion, we decreased its weight from 100% to 0% in increments of 10%, redistributing the reduced weight among the remaining criteria. This procedure was applied to each criterion, maintaining the constraint that the cumulative weight equals 1. Subsequently, we iterated the IVPF-WASPAS method using these adjusted criterion weights. The impact of these weight adjustments on the ranking performance of alternatives for each decision maker is depicted in Figures 5, 6 and 7 respectively.

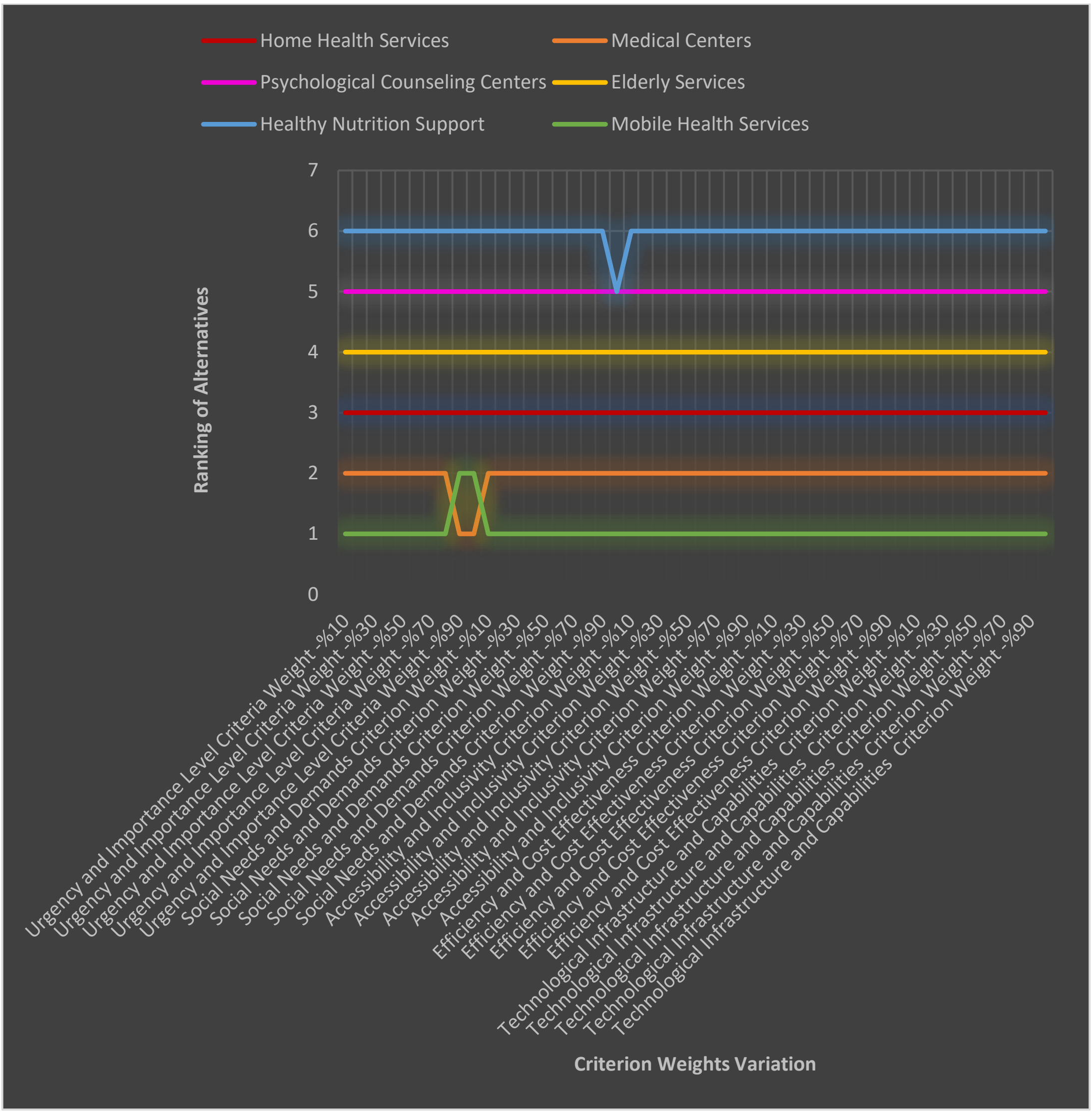


Figure 5. Effects Of Changes in Criterion Weights on The Ranking Performance of Alternatives for Decision Maker-1

Examining Figure 5 reveals the following results:

Reducing the weight of the "Urgency and Importance Level" criterion by 90% and 100% causes the "Mobile Health Services" alternative, initially in the first position, to fall to the second position.

Conversely, the "Medical Centers" alternative, which is initially in the second position, rises to the first position when the weight of the "Urgency and Importance Level" criterion is reduced by 90% and 100%.

The rankings of the "Home Health Services" in the third position, "Elderly Services" in the fourth position, and "Psychological Counseling Centers" in the fifth position remain unchanged regardless of any variations in criterion weights.

The "Healthy Nutrition Support" alternative, which is initially in the sixth position, moves to the fifth position when the weight of the "Social Needs and Demands" criterion is reduced by 100%.

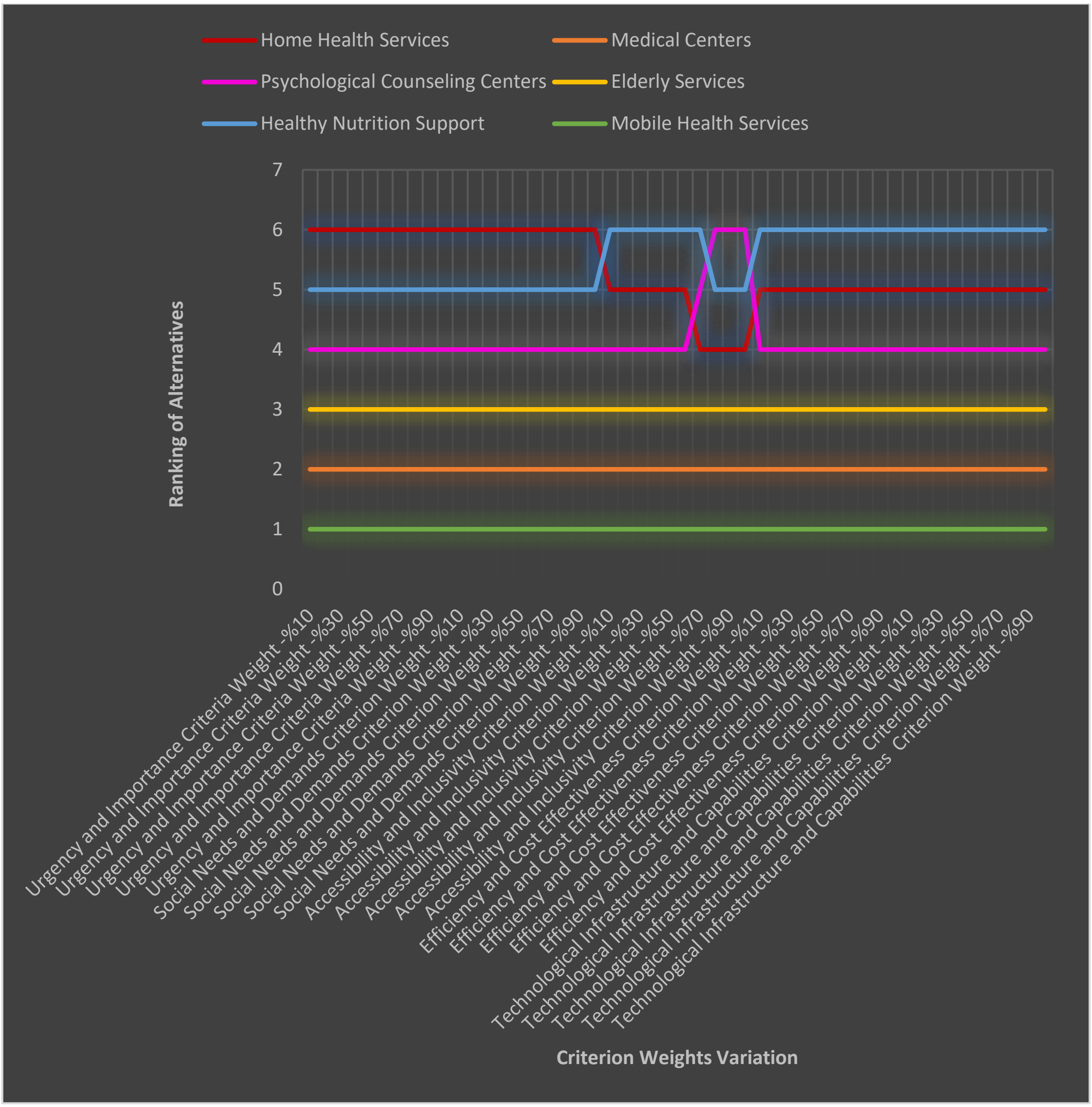


Figure 6. Effects Of Changes in Criterion Weights On The Ranking Performance Of Alternatives For Decision Maker-2

Examining Figure 6 reveals the following results:

The rankings of the alternatives "Mobile Health Services" in the first position, "Medical Centers" in the second position and "Elderly Services" in the third position did not change with any variation in criterion weights.

Reducing the "Accessibility and Inclusivity" criterion weight by 70% moves the "Psychological Counseling Centers" alternative from the fourth to the fifth position, and reducing this criterion weight by 80%, 90%, and 100% moves it to the sixth position.

The "Healthy Nutrition Support" alternative, which is in the fifth position, moved to the sixth position when the criterion weight of "Accessibility and Inclusivity" was reduced by 10%, 20%, 30%, 40%, 50%, 60% and 70%, as well as when the criterion weight of "Efficiency and Cost Effectiveness" was reduced by 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100%, and when the criterion weight of "Technological Infrastructure and Capabilities" was reduced by 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100%.

The "Home Health Services" alternative, which is in the sixth position, moved to the fifth position when the criterion weight of "Accessibility and Inclusivity" was reduced by 10%, 20%, 30%, 40%, 50% and 60%, and when the criterion weight of "Efficiency and Cost Effectiveness" was reduced by 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100%, and when the criterion weight of "Technological Infrastructure and Capabilities" was reduced by 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100%; it moved to the fourth position when the criterion weight of "Accessibility and Inclusivity" was reduced by 70%, 80%, 90% and 100%.

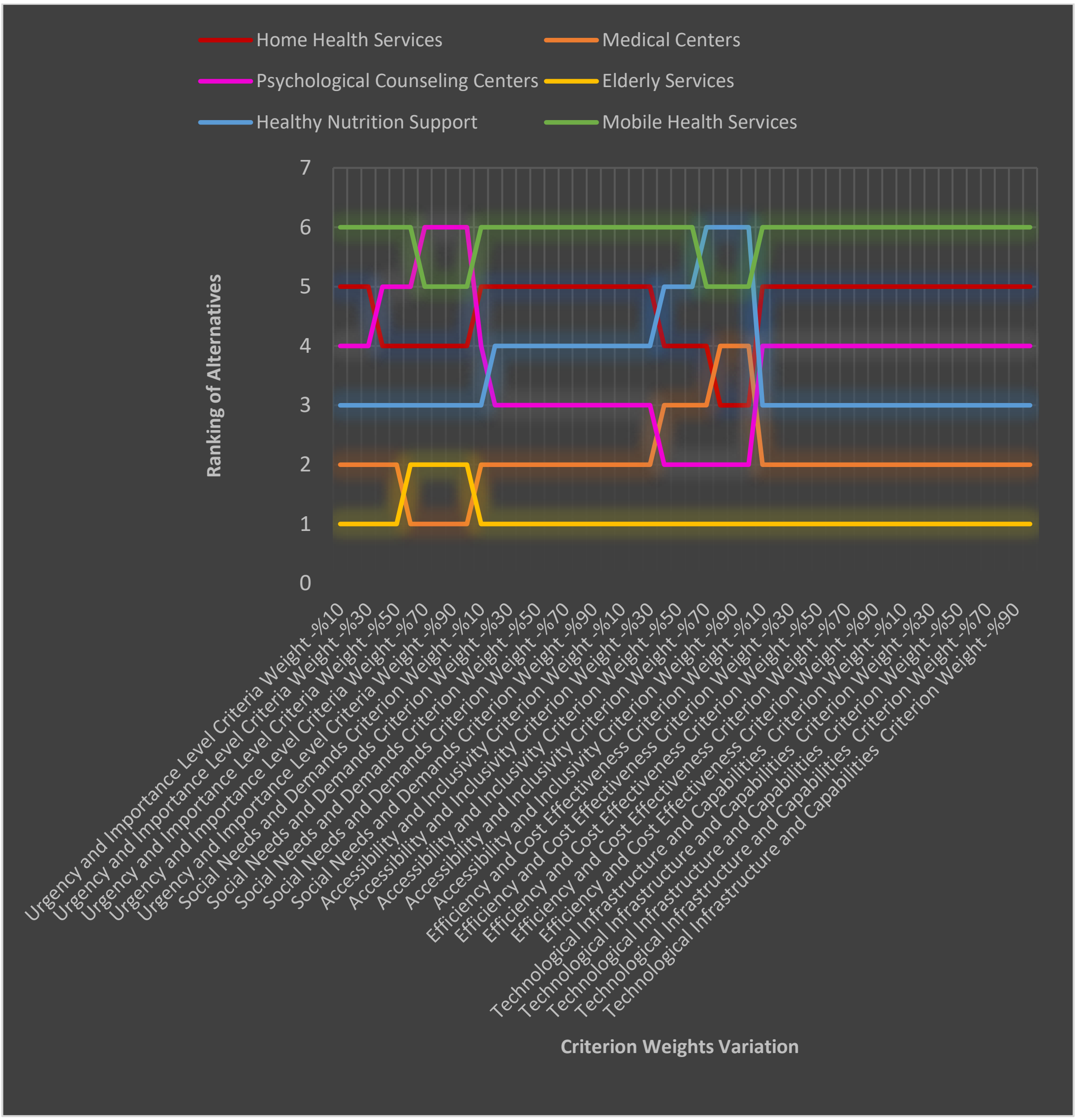


Figure 7. Effects of Changes in Criterion Weights on The Ranking Performance of Alternatives for Decision Maker-3

Examining Figure 7 reveals the following results:

The "Elderly Services" alternative, which was ranked first, dropped to the second position when the "Urgency and Importance Level" criterion weight was reduced by 60%, 70%, 80%, 90% and 100%.

The "Medical Centers" alternative, which was ranked second, rose to the first position when the "Urgency and Importance Level" criterion weight was reduced by 60%, 70%, 80% and 90%; dropped to the third position when the "Accessibility and Inclusivity" criterion weight was reduced by 40%, 50%, 60% and 70%; and further dropped to the fourth position when the "Accessibility and Inclusivity" criterion weight was reduced by 80%, 90% and 100%.

The "Healthy Nutrition Support" alternative, which was ranked third, dropped to the fourth position when the "Social Needs and Demands" criterion weight was reduced by 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100%, and when the "Accessibility and Inclusivity" criterion weight was reduced by 10%, 20% and 30%; dropped to the fifth position when the "Accessibility and Inclusivity" criterion weight was reduced by 40%, 50% and 60%; and further dropped to the sixth position when the "Accessibility and Inclusivity" criterion weight was reduced by 70%, 80%, 90% and 100%.

The "Psychological Counseling Centers" alternative, which was ranked fourth, dropped to the fifth position when the "Urgency and Importance Level" criterion weight was reduced by 40%, 50% and 60%; further dropped to the sixth position when the "Urgency and Importance Level" criterion weight was reduced by 70%, 80%, 90% and 100%; rose to the third position when the "Social Needs and Demands" criterion weight was reduced by 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100%, and when the "Accessibility and Inclusivity" criterion weight was reduced by 10%, 20% and 30%; and further rose to the second position when the "Accessibility and Inclusivity" criterion weight was reduced by 40%, 50%, 60%, 70%, 80%, 90% and 100%.

The "Home Health Services" alternative, which was ranked fifth, rose to the fourth position when the "Urgency and Importance Level" criterion weight was reduced by 40%, 50%, 60%, 70%, 80%, 90% and 100%, and when the "Accessibility and Inclusivity" criterion weight was reduced by 40%, 50%, 60% and 70%; and further rose to the third position when the "Accessibility and Inclusivity" criterion weight was reduced by 80%, 90% and 100%.

The "Mobile Health Services" alternative, which was ranked sixth, rose to the fifth position when the "Urgency and Importance Level" criterion weight was reduced by 70%, 80%, 90% and 100%, and when the "Accessibility and Inclusivity" criterion weight was reduced by 70%, 80%, 90% and 100%.

Overall evaluation of the sensitivity analysis results shows minor changes in the preference rankings of alternatives due to modifications in criterion weights, indicating that these changes do not significantly alter the general outcomes. Despite modifications in criterion weights, the results obtained demonstrate that the proposed integrated approach is a robust, accurate, realistic, and reasonable technique that yields strong outcomes.

6. Conclusion and Discussion

In this research, three decision-makers evaluated six alternatives within the framework of five criteria: "Urgency and Importance Level", "Social Needs and Demands", "Accessibility and Inclusivity", "Efficiency and Cost Effectiveness" and "Technological Infrastructure and Capabilities". The PF-CRITIC method determined the weights of the selection criteria. Subsequently, we inputted the obtained criterion weights into the IVPF-WASPAS method to rank the alternatives. The conducted study demonstrated the feasibility of the proposed framework.

The decision-making process involved aggregating different criteria or preferences to make an overall assessment. According to the PF-CRITIC method, decision maker-1 prioritized "Social Needs and Demands" as the most important criterion, while decision makers 2 and 3 prioritized "Accessibility and Inclusivity". In contrast, the IVPF-WASPAS method identified "Mobile Health Services" as the most important alternative in the shared importance ranking, with "Healthy Nutrition Support" deemed the least important.

We conducted a comprehensive sensitivity analysis based on variations in criterion weights to test the validity of the proposed integrated model. The results indicate that the proposed integrated approach is a robust, accurate, reasonable, and realistic technique that yields strong outcomes.

We believe that the integrated method used in the study contributes by providing practitioners with a methodological framework, thereby offering comprehensive insights into the study and its methods, which can guide researchers intending to undertake similar studies. The proposed hybrid approach can also be applied to solve decision-making problems encountered in various fields. The presented framework can serve as an exemplary model and lay the groundwork for future research. Below, we delineate the limitations of the study.

To obtain reasonable and realistic results, researchers should carefully select experts. In the coming years, those conducting research in this field may benefit from collaborating with highly knowledgeable, experienced, and authoritative experts, as demonstrated in the study. Moreover, while selecting the correct and appropriate criteria is crucial, relying solely on a literature review may not suffice for determining these criteria. Therefore, conducting fieldwork in collaboration with experts, as practiced in this study, can help define suitable selection criteria.

Future research could employ the methodology used in this document to address issues through diverse evaluations using various FMCDM methods. Additionally, expanding the number of experts in the decision-making group can enhance the robustness of outcomes.

References

[1] Buchert, U., Kemppainen, L., Olakivi, A., Wrede, S., & Kouvonen, A. (2022). Is digitalisation of public health and social welfare services reinforcing social exclusion? the case of russian-speaking older migrants in finland. Critical Social Policy, 43(3).

[2] Schou, J. and Pors, A. S. (2018). Digital by default? a qualitative study of exclusion in digitalised welfare. Social Policy &Amp; Administration, 53(3), 464-477. https://doi.org/10.1111/spol.12470.

[3] Lloyd, C. and Payne, J. (2021). Fewer jobs, better jobs? an international comparative study of robots and 'routine' work in the public sector. Industrial Relations Journal, 52(2), 109-124. https://doi.org/10.1111/irj.12323.

[4] Collington, R. (2021). Disrupting the welfare state? digitalisation and the retrenchment of public sector capacity. New Political Economy, 27(2), 312-328. https://doi.org/10.1080/13563467.2021.1952559.

[5] Tiainen, M., Ahonen, O., Hinkkanen, L., Rajalahti, E., & Värri, A. (2021). The definitions of health care and social welfare informatics competencies. Finnish Journal of eHealth and eWelfare, 13(2), 147-159.

[6] Baumgartner, M., Sauer, C., Blagec, K., & Dorffner, G. (2022). Digital health understanding and preparedness of medical students: a cross-sectional study. Medical education online, 27(1), 2114851.

[7] Gopal, G. V., Suter-Crazzolara, C., Toldo, L., & Eberhardt, W. (2018). Digital transformation in healthcare – architectures of present and future information technologies. Clinical Chemistry and Laboratory Medicine (CCLM), 57(3), 328-335.

[8] Scarano, G. and Colfer, B. (2022). Linking active labour market policies to digitalisation–a review between remote and automated possibilities. International Journal of Sociology and Social Policy, 42(13/14), 98-112.

[9] Holm, S. G., Mathisen, T. A., Sæterstrand, T. M., & Brinchmann, B. S. (2017). Allocation of home care services by municipalities in norway: a document analysis. BMC Health Services Research, 17(1). https://doi.org/10.1186/s12913-017-2623-3.

[10] Shava, E. and Vyas-Doorgapersad, S. (2022). Fostering digital innovations to accelerate service delivery in south african local government. International Journal of Research in Business and Social Science (2147- 4478), 11(2), 83-91.

[11] Yager, R. R. (2013, June). Pythagorean fuzzy subsets. In 2013 joint IFSA world congress and NAFIPS annual meeting (IFSA/NAFIPS) (pp. 57-61). IEEE.".

[12] Atanassov, K. T. (1986). Intuitionistic fuzzy sets. Fuzzy Sets and Systems, 20(1), 87-96. https://doi.org/10.1016/S0165-0114(86)80034-3.

[13] Wang, J. C., & Chen, T. Y. (2018). Multiple criteria decision analysis using correlation-based precedence indices within Pythagorean fuzzy uncertain environments. International Journal of Computational Intelligence Systems, 11(1), 911-924.

[14] Peng, X., & Selvachandran, G. (2019). Pythagorean fuzzy set: state of the art and future directions. Artificial Intelligence Review, 52(3), 1873–1927. https://doi.org/10.1007/s10462-017-9596-9.

[15] Zhang, X., & Xu, Z. (2014). Extension of TOPSIS to multiple criteria decision making with Pythagorean fuzzy sets. International journal of intelligent systems, 29(12), 1061-1078.

[16] Peng, X., & Yang, Y. (2016). Fundamental properties of interval-valued Pythagorean fuzzy aggregation operators. International Journal of Intelligent Systems, 31(5), 444-487.

[17] Sancar, S. (2022). Pisagor bulanık AHP ve pisagor bulanık WASPAS yöntemleri ile bakım stratejisi seçimi: Gazete matbaası örneği (Master's thesis, İbn Haldun Üniversitesi, Lisansüstü Eğitim Enstitüsü).

[18] Peng, X., Zhang, X., & Luo, Z. (2020). Pythagorean fuzzy MCDM method based on CoCoSo and CRITIC with score function for 5G industry evaluation. Artificial Intelligence Review, 53(5), 3813-3847. https://doi.org/10.1007/s10462-019-09780-x.

[19] Kabak, M., & Erdebilli, B. (2021). Bulanık Çok Kriterli Karar Verme Yöntemleri - MS Excel ve Software Çözümlü Uygulamalar. Ankara: Nobel Yayın.

[20] Turskis, Z., Zavadskas, E. K., Antucheviciene, J., & Kosareva, N. (2015). A hybrid model based on fuzzy AHP and fuzzy WASPAS for construction site selection. International Journal of Computers communications & control, 10(6), 113-128.

[21] Ilbahar, E., & Kahraman, C. (2018). Retail store performance measurement using a novel interval-valued Pythagorean fuzzy WASPAS method. Journal of Intelligent & Fuzzy Systems, 35(3), 3835-3846.

[22] Görçün, O. F., Senthil, S., & Küçükönder, H. (2021). Evaluation of tanker vehicle selection using a novel hybrid fuzzy MCDM technique. *Decision Making: Applications in Management and Engineering*, *4*(2), 140-162.

[23] Pérez-Domínguez, L., Rodríguez-Picón, L. A., Alvarado-Iniesta, A., Luviano Cruz, D., & Xu, Z. (2018). MOORA under Pythagorean fuzzy set for multiple criteria decision making. Complexity, 2018. https://doi.org/10.1155/2018/2602376

[24] O'cathain, A., Murphy, E., & Nicholl, J. (2008). The quality of mixed methods studies in health services research. *Journal of health services research & policy*, *13*(2), 92-98.

[25] Haktanir, E., & Kahraman, C. (2022). A novel picture fuzzy CRITIC & REGIME methodology: Wearable health technology application. *Engineering Applications of Artificial Intelligence*, *113*, 104942.

[26] Wang, Y., Wang, W., Wang, Z., Deveci, M., Roy, S. K., & Kadry, S. (2024). Selection of sustainable food suppliers using the Pythagorean fuzzy CRITIC-MARCOS method. *Information Sciences*, *664*, 120326.

[27] Gedikli, T., & Cayir Ervural, B. (2022). Identification of optimum COVID-19 vaccine distribution strategy under integrated Pythagorean fuzzy environment. In *Digitizing Production Systems: Selected Papers from ISPR2021, October 07-09, 2021 Online, Turkey* (pp. 65-76). Springer International Publishing.

■ RESEARCH ARTICLE

Sentiment Analysis in Turkish Tweets Using Different Machine Learning Algorithms

Hunaida Avvad [a†], **Ecem Ereren** [a]

[a] Department of Management Information Systems, İzmir Bakırçay University, İzmir, Türkiye
† hunaida.awwad@bakircay.edu.tr, corresponding author

RECEIVED AUGUST 12, 2024
ACCEPTED SEPTEMBER 19, 2024

CITATION Avvad, H., & Ereren, E. (2024). Sentiment analysis in Turkish tweets using different machine learning algorithms. *Artificial Intelligence Theory and Applications, 4(*2), 107-120.

Abstract

Understanding emotions in any written text is considered a hot topic for many researchers in the field of text mining, especially with the large contribution of users over the web 2.0 and with the growth of the different social media platforms. In this study, we analysed emotions in Turkish text and studied the sentiment within each document using sentiment analysis techniques. Sentiment analysis is the process of identifying and evaluating the emotional states contained in texts. This study aimed to investigate the effect and accuracy rate of sentiment analysis in Turkish texts. Sentiment analysis is an important field of research that helps to obtain important data in many areas, such as marketing, social media analysis, and customer feedback. A comprehensive data set consisting of Turkish tweets from Kaggle was used, and the emotional states of the texts were labelled. This data set consists of a variety of tweets with different topics and emotional tones. Using natural language processing techniques and machine learning algorithms, the data set was processed, and the model was trained. Within the scope of the study, different root extraction methods and a vector space model were used. In addition, machine learning algorithms such as Naive Bayes, Random Forest, Decision Tree, Gradient Boosting, Bernoulli Naive Bayes, Logistic Regression, K-Neighbours-Classifier, and Support Vector Classifier were applied to evaluate accuracy. This study aims to emphasize the importance of sentiment analysis in Turkish texts, examine the impact of the methods used, and form a basis for future studies.

Keywords: sentiment analysis, Turkish text, machine learning, Turkish tweet

1. Introduction

Social media platforms and Web 2.0 allowed people to share their experience and to express their feedback about many products/services that they received, the huge size of the written text on the Web 2.0 is considered a hot research topic for many researchers who focus on text mining in order to analyse emotions in any written text is considered a hot topic for many researchers in the field of text mining. Social media tools such as Twitter and Facebook have an important role to play as big data sources in the process of extracting information from any text. The most important reason for this is that the text data produced by these applications is increasing significantly day by day. Sentiment Analysis (SA) has emerged as a field in which natural language processing, machine learning, and linguistic methods are used to understand the emotional tone of texts and

Artificial Intelligence Theory and Applications, ISSN: 2757-9778. ISBN: 978-605-69730-2-4 © 2024 İzmir Bakırçay University

identify emotional trends or moods in the text. SA can be applied in many areas, such as social media analytics, customer feedback, marketing strategies, product reviews, and survey responses. For example, a company may try to understand customer satisfaction and perception by analysing tweets about their brand or products on social media platforms. Similarly, a movie studio can gauge the overall emotional response of movies by analysing the comments that viewers share about the movies on social media. It is of great importance to analyse the big data, which has emerged with the increase in the use of the internet and social media, which has become widespread today, and to transform it into meaningful information. SA is the process of systematically examining the data containing opinions in a text and determining the emotion category and emotion polarity of the text. SA approaches are frequently used not only in linguistics, but also in many different fields such as financial markets, marketing, and social media analysis, Tokcaer [1]. In SA studies, it is questioned and analysed whether the texts have positive, negative, or neutral content. According to the results of this analysis, the attitude of individuals or a certain group about the subject related to the study is determined. In this respect, SA can guide businesses on issues such as preliminary market research for a new product to be launched, how a decision to be taken for a community will receive a positive or negative reaction, and whether people who will watch the movie decide to watch the movie according to previous comments. However, the large amount of data from which a positive or negative opinion can be obtained makes it difficult to make this analysis by examining it one by one. Therefore, sentiment analysis has become one of the most important and studied topics in the fields of text mining and machine learning, Kaynar et al. [2]. Danisman and Alpkocak [3] used different machine learning algorithms such as Vector Space Model, Naïve Bayes, and SVM classifiers to compare between the performances using the ISEAR data set that contains 5 classes of emotions: anger, disgust, fear, joy, and sadness. The training set is enriched with WordNet Affect and WPARD (Wisconsin Perceptual Attribute Rating Database) data sources, Medler et al. [4]. Stop word removal and root removal operations were applied, and the term frequency – inverse document frequency (TF-IDF) method was chosen as feature weighting. According to the results obtained, an overall classification accuracy of 70.2% was achieved. There are many studies on sentiment analysis on text data sources, especially in English. Sentiment analysis generally uses two basic approaches: the rule-based approach and the machine learning-based approach, Alpkoçak et al. [5]. Rule-based approaches use predefined rules that include certain emotional words or phrases that indicate a particular emotion. Machine learning-based approaches, on the other hand, use large data sets to automatically learn how emotional expressions relate to specific emotions and create classification models. Identifying and classifying emotional expressions in Turkish tweets is an important issue among text classification problems. Social media platforms, which are easily accessible platforms, provide remarkable resources to get feedback from target audiences, but it is impossible to analyse these feedbacks with human labour. Therefore, automated sentiment analysis tools are essential for companies' customer service to be able to capture complaints and/or positive feedback at the right time. Processing by computer allows this data to be used in the market, Türkmenoğlu [6].

Studies such as Kozareva et al. [7], Mohammad [8], and Chaffar and Inkpen [9] have carried out important contributions on sentiment analysis. In the Turkish language, there are limited number of studies on the subject of sentiment analysis, Boynukalın [10]. In the Turkish language, there are many to be done because of the lack of studies dealing with SA different aspects and subdomains. The Turkish language is known as a sticky language. With the use of derived suffixes, the root of a word can be transformed into a completely different type of word, for example, from a noun to a verb. These derivatives can be applied consecutively more than once. Since each derived suffix has the potential to change the meaning of a word, each derived suffix must be examined separately to obtain the true meaning of a word. Previous studies have often focused on official data sources, such as newspaper headlines and surveys. Recently, however, research on informal data sources such as instant messaging, blog posts, and Twitter has become popular. Twitter is a social micro-blogging service that allows users to publish and read messages in real time, and these messages are called "tweets". People share their thoughts, daily life events, and feelings on Twitter. Although there are many micro-blogging platforms, Twitter is the most popular. The large volume of user-generated content makes Twitter a suitable space for sentiment analysis. The similarity of the tweets also makes them effectively actionable for sentiment detection tasks. Boynukalın [10] used a translation of the ISEAR data set and a manually marked data set to classify the Turkish texts. Apart from emotion classes, the determination of emotion levels was also attempted. Different combinations of n-gram features were used. A weighted log probability algorithm was used to score the features and identify the most important ones. Kaya et al. [11] applied supervised classification algorithms for the sentiment classes positive and negative in Turkish news columns. With the exception of SVM, Maximum Entropy, and Naïve Bayes classifiers, the character-based n-gram language model was used. This language model uses characters instead of words as a unit. Their idea is that statistical methods may not yield promising results due to the fact that Turkish is a morphologically rich language.

In this study, we analysed emotions in Turkish text and studied the sentiment within each document using Sentiment Analysis (SA) techniques. This article aims to show the different methods and approaches that can be used to evaluate sentiment analysis and classify emotional expressions in Turkish texts. This study was carried out in order to evaluate the effectiveness of existing methods and algorithms for sentiment analysis in Turkish and to propose new approaches to obtain better results for sentiment analysis in Turkish texts. In this study, a ready-made Turkish tweet data set downloaded from the Kaggle platform was used. The data set consists of Turkish tweets shared by different topics and users. The distribution of classes in the used data set was not balanced. For imbalanced data sets, some classification algorithms may perform better; for example, Decision Trees, Random Forests, or Gradient Boosting models have been used in this data set because they can work well in unbalanced data sets. We also split our data set into training and testing data, making it available for training and testing our machine learning models. After performing the data preprocessing steps, we used the processed data for building the ML models, testing them, and evaluating them one by one.

After carrying out the needed preprocessing steps, a classification model was developed to conduct sentiment analysis using various machine learning algorithms. These algorithms include popular methods such as Support Vector Machines (SVM), Decision Trees (DT), and Random forests (RF), in addition to, Logistic Regression (LR), K-nearest Neighbours (KNN), and Gradient Boosting (GB). The developed models were used to classify emotional expressions in Turkish tweets as positive, negative, or neutral. During the training process, the data set was split, and the performance of the models was evaluated using the five-fold cross-validation method. Performance metrics such as accuracy rates and confusion matrices were calculated, and the results were analysed. In addition, the in-class performance values of the model were calculated.

2. Literature review

In recent years, many studies have been carried out on sentiment analysis in Turkish texts. Especially considering that most of the social media data language used in Turkey is expressed using Turkish language, that leads to the increase of the importance of sentiment analysis in Turkish texts. In many studies, various approaches have been adopted to detect and classify emotional expressions in Turkish texts by using different methods such as machine learning algorithms, natural language processing techniques, and deep learning models. The researchers evaluated the results obtained using performance metrics such as accuracy, precision, recall and F1 score, and proposed new methods and improvements to increase the success of sentiment analysis in Turkish texts.

There are many studies on SA, especially in English, on text data sources. Researchers such as Kozareva et al. [7], Mohammad [8] and Chaffar and Inkpen [9] have done important studies on this subject. In Turkish, there are fewer studies on the subject of SA, Boynukalın [10]. In the Turkish language, more work has been done on Natural Language Processing (NLP) rather than sentiment analysis because NLP field is more developed. Danisman and Alpkocak [3] compared the performances of Vector Space Model, Naïve Bayes and SVM classifiers using ISEAR dataset for 5 emotion classes: anger, disgust, fear, joy, and sadness. The training set is enriched with WordNet Affect and WPARD (Wisconsin Perceptual Attribute Rating Database) data sources, Medler et al. [4]. Stop word removal and root removal operations were applied, and the TF-IDF method was chosen as feature weighting. According to the results obtained, an overall classification accuracy of 70.2% was achieved.

The Turkish language is known as a sticky language. Using derived suffixes, the root of a word can be transformed into a completely different type of word, for example, from a noun to a verb. These derivations can be applied sequentially more than once, Oflazer [12]. Since each derived suffix has the potential to change the meaning of the word, each derived suffix must be examined separately to get the true meaning of a word. Previous research has often focused on official data sources such as newspaper headlines and surveys. Recently, however, research on informal data sources such as instant messaging, Neviarouskaya [13], blog posts, Wang [14] and Twitter, Mohammad [8] has become popular. Twitter is a social micro-blogging service that allows users to post and read messages in real time, and these messages are called "tweets". People share their thoughts, daily life events and feelings on Twitter. Although there are many micro-blogging platforms, Twitter is the most popular. The sheer volume of user-generated content makes Twitter a viable space for sentiment analysis. The similarity of tweets also makes them effectively workable for emotion detection tasks. Boynukalın [10] used a translation of the ISEAR dataset and a manually marked dataset to classify Turkish texts. Apart from emotion classes, determination of emotion levels was also attempted. Different combinations of n-gram features were used.

Kaya et al. [11] applied supervised classification algorithms for positive and negative emotion classes in Turkish news columns. Except for SVM, Maximum Entropy\, and Naïve Bayes classifiers, the character-based n-gram Language Model was used. This language model uses characters instead of words as units. Their thoughts are that statistical methods may not yield promising results because Turkish is a morphologically rich language. Erogul [15] created a dataset from a Turkish movie review site. Reviews were labelled by their authors with positive, negative, or neutral icons. In the generated dataset, an emotionally labelled data item was created by using the text and symbol of the review together. A polarity dataset was created from another movie review site, which includes the ratings given to the movies by the users. Combinations of n-grams and POS information for the classification task were used for morphological analysis using the Zemberek tool. For the polarity labelled dataset, scores were estimated using regression and single comparison techniques.

3. Methodology

We used Python with the data set downloaded from Kaggle to detect and evaluate different ML algorithms. In this study, various ML algorithms have been adopted to detect and classify emotional expressions in Turkish texts after applying the data pre-processing steps. We evaluated the results obtained using performance metrics such as accuracy, precision, recall, and F1 score, and proposed new methods and improvements to increase the success of sentiment analysis in Turkish texts.

3.1. Data Collection and Pre-Processing

In this study, we used a Turkish tweet data set from the Kaggle platform to perform sentiment analysis in Turkish texts. The data set consists of a variety of tweets with different topics and emotional tones. We obtained this data set, which included a total of 4201 tweets, and labelled the emotional states.

In the data pre-processing phase, we removed unnecessary characters, special symbols, and punctuation marks from the texts. In addition, we converted the texts to lowercase and performed stemming by using Turkish language processing libraries. Thus, we brought it to a simpler format without affecting the meaning of the texts and purified it from unnecessary information. Stemming is a text processing technique used in Natural Language Processing (NLP) and Computer Language Processing (CLP). Basically, it aims to extract word roots, or the basic form of the word. This is used to identify similarity between different variations or trends of a word and make text analysis or information extraction easier and more effective.

3.2. Text Representation

In order for the texts to be processed with ML algorithms, text must be converted to a vector format. There are three different ways that can help in text representations; bag of words (BOW), n-grams, and Term Frequency-Inverse Document Frequency (TF-IDF). In the bag of words, each word in a document will be added to a bag without any repetition and without keeping the sequence of words existence in the document. A matrix of 1Xn in which n represents the number of words in the bag will be used with a word frequency; words with higher occurrences show that they are more common in the document. One drawback of BOW that words in stop words list are included in the bag without removing them, and they have the highest frequencies, which will affect the vectors negatively. Another drawback of BOW is that it can’t perform well when you have similar documents with small changes. In the n- gram method, the grams (words) will be treated as 2- gram or commonly called bigram, so the model will check the frequency of words in a document as pairs. N-gram can keep the relationship between the consecutive words better than BOW but because of data sparsity n-gram can fail in building a good model specially with there is low frequencies of the n-grams.

For the aforementioned drawbacks of both BOW and n-grams, we applied text representation methods such as TF-IDF to provide vector representation of Turkish texts. In particular, by obtaining TF-IDF word vectors, we aimed to better capture the meaning and emotional content in the texts. TF-IDF is a text mining technique used in NLP applications such as text mining, text classification, and information extraction. TF-IDF is used to determine the importance of a term within a given document and to compare the importance of those terms within a given collection of documents. The main purpose of TF-IDF is to determine the weight and importance of terms between textual documents. TF-IDF increases the applicability of text mining algorithms.

We applied the TF-IDF technique to the processed text after the pre-processing phase and used it as a feature extraction method to convert text documents into numerical data. Term frequency is calculated by finding the number of occurrences of each term (word) in a document, then it will be multiplied by the inverse document frequency, which represents how common a word is in the corpus.

3.3. Machine Learning Algorithms

After completing the data pre-processing and text representation steps, we performed SA by using different machine learning algorithms in the training and evaluation processes. Within the scope of our experiments, we evaluated popular classification algorithms such as Random Forest (RF), Logistic Regression (LG), K-Nearest Neighbours (KNN), Bernoulli Naive Bayes (BNB), Decision Tree (DT), and Support Vector Classifier (SVC).

Random Forest (RF): RF is an algorithm that produces and classifies multiple decision trees by training each one on a different observation sample. The algorithm creates a decision tree for each sample, and the estimated value result of each decision tree is formed. Voting is performed for each value formed as a result of the prediction. Observation is assigned to the class with the most votes.

Logistic Regression (LR): LR is a supervised ML classification algorithm that aims to predict the probability that an instance belongs to a given class or not. Then the data point will be assigned to the class with the highest probability.

K-Nearest Neighbours (KNN) Classifier: KNN classifier classifies using similar samples around labelled data points, KNN is based on deciding the class of the data point depending on the class that is nearest neighbours of the vector. K here represents how many neighbour points we are going to check. The distance will be calculated between the data point of that we want to assign its class and the K nearest points.

Bernoulli Naive Bayes (BNB) Classifier: NB classifier is a probability-based classification algorithm based on Bayes' theorem that makes use of Bayes Theorem during the training phase.

Gradient Boosting (GB): is a famous boosting algorithm using ensemble learning methods that enhance the results of training model sequentially that each model will enhance the previous one.

Decision Tree (DT): DT is one of the tree-based learning algorithms. It is a tree structure that performs classification by dividing the data set according to its characteristics.

Support Vector Classifier (SVC): SVC classifier classify with the supervised learning method. It is considered as a powerful classification method that attempts to find a hyperplane with a maximum margin between different classes. It aims to have this line at the maximum distance for the points of both classes.

3.4. Model Performance and Evaluation

We performed model training and evaluation for each algorithm with the five-fold cross-validation method. Thus, we have increased the reliability of model performance and prevented problems such as overfitting. Using confusion matrix, we evaluated number of evaluation metrics such as accuracy, precision, recall and F1 score. By comparing our results, we tried to identify the most effective and successful algorithms. Using these methods, we aim to achieve successful results in SA in Turkish texts. We apply these methods to understand the effectiveness of different ML algorithms in SA and to classify emotional content in Turkish texts more accurately and effectively.

4. Experiments and Results

In this study, we evaluated different ML algorithms to perform SA in Turkish texts, and the results we obtained were quite remarkable. Experiments conducted on various tweets in our data set provided important insights into the effectiveness and accuracy rates of the algorithms used for SA.

```
# Metin verilerini özelliklere dönüştürme
X_features = vectorizer.transform(df['Stemmed_Tweets'])

# Sentiment değerlerini hesaplama
sentiment_values = model.predict(X_features)

# Hesaplanan sentiment değerlerini veri setine ekleme
df['Sentiment'] = sentiment_values


# Hesaplanan sentiment değerlerini veri setine ekleme
df['Sentiment'] = sentiment_values

# Veri setinin ilk 10 gözlemi ve Sentiment sütununu görüntüleme
print(df.head(10)['Sentiment'])

0    1
1    1
2    1
3    1
4    0
5    1
6    1
7    1
8    1
9    1
Name: Sentiment, dtype: int64
```

Figure 1. Calculation of Sentiment Values of the Data

The code in Figure1 uses a vectorizer to convert text data from a data set into features and calculates sentiment values using these features. It then adds the calculated sentiment values to the df data set and then prints the first 10 observations of the data set and the "Sentiment" column on the screen.

```
# Turizm kelimesini içeren tweetleri seçme
turizm_tweets = df[df['Stemmed_Tweets'].str.contains('turizm', case=False)]

# Seçilen tweetlerin sentiment değerlerini hesaplama
sentiment_values = model.predict(vectorizer.transform(turizm_tweets['Stemmed_Tweets']))

# Hesaplanan sentiment değerlerini veri setine ekleme
turizm_tweets['Sentiment'] = sentiment_values

# Turizm tweetlerini ve sentiment değerlerini görüntüleme
print(turizm_tweets[['Stemmed_Tweets', 'Sentiment']])

                                         Stemmed_Tweets  Sentiment
0      say cumhurbaşka turizm ertelendik biliyor faka...          1
3      sınav tarih değiştirir pedalogu psikologu danı...          1
4                   turizm yıl gençlik gelecek kurtarır          0
5      siz kıyak isteye yok turizm yüz yedik hakk ger...          1
6      p söylemek bil üzüyor turizm kadar değerli sın...          1
...                                                 ...        ...
4189   ttga nın başarıl çalışma ülke turizm değerli k...          2
4193   temel olarak konaklama içeriyor fakat turizm g...          0
4194                       avrupal turizmci ortak test pla          0
4195               turizmle ilgil kalem fiyat kadar yüksel          1
4198                     turizm ye bir hikaye ihtiyaç var          1
```

Figure 2. Sentiment Values of the Data Containing the Word ‘turizm’

The code in Figure 2 selects tweets that contain the word tourism; the number of tweets that contain the word ‘turizm’ is 840, calculates the sentiment values of those tweets, adds these values to the tourism tweets sub data set, and finally displays the tourism tweets and sentiment values. The experiments that have been included in the study are applied to the sub data set of size 840.

The confusion matrix allows us to evaluate the performance of the classification model in more detail. However, based solely on the results of this matrix, it is difficult to determine with certainty how good the performance of the model is. The confusion matrix can be used to understand how the model performs in certain classes, but it needs to be considered in conjunction with other performance metrics in order to fully evaluate performance. To evaluate the confusion matrices in more details, we calculated the performance metrics of each class, such as precision, recall, and F1 score. We can also evaluate these metrics on a class-by-class basis to understand the performance differences between classes.

According to the results of the experiments shown in Table1, Logistic Regression (LR) and Support Vector Classifier (SVC) were found to have the highest accuracy rates of SA in Turkish texts (0.62). Random Forest (RF) and Bernoulli Naïve Bayes (BNB) accuracies are almost similar to Logistic Regression (LR) and Support Vector Classifier (SVC) with accuracy rate of 0.61. However, the K-Nearest Neighbours (KNN) algorithm achieved a slightly lower accuracy rate with 0.48 compared to other methods. To be able to decide which algorithm(s) work better than the others, a confusion matrix analysis was performed for all experiments. Table2 shows the confusion matrices for RF, LR, KNN, BNB, GB, DT, and SVC respectively. The confusion matrix here is a multi-class model of size 3X3, Negative class, Positive class, and Neutral class. Using the confusion matrix for each experiment we calculated Accuracy, Precision, Recall, and F1-score evaluation metrics for each class as shown in Table 3. From Table 3 we can see that the best achieved result is for BNB’s Neutral class with 0.75 accuracy rate followed by GB’s Neutral class with 0.74 accuracy rate. For the Positive class, the best achieved accuracy is for SVC with 0.74 followed by LR with 0.71. For the Negative class, LR and DT have the highest accuracy rate with 0.54 and 0.45, respectively.

F1- scores measures the harmonic mean of the Precision and Recall, Table3 shows that the F1-score for Positive mood class is higher than the F1-score for both the Negative and the Neutral classes for all experiments except KNN. The best achieved F1-score is for BNB’s Positive class with 0.72, followed by SVC’s Positive class with 0.71.

Table 1. Summary Table

Experiments	**Accuracy**	**Precision**	**Recall**	**F1 score**
Random Forest (RF)	0.61	0.62	0.61	0.61
Logistic Regression (LR)	0.62	0.62	0.62	0.62
K-nearest Neighbors (KNN)	0.48	0.48	0.48	0.48
Bernoulli Naïve Bayes (BNB)	0.61	0.62	0.61	0.61
Gradient Boosting (GB)	0.55	0.58	0.55	0.53
Decision Tree (DT)	0.57	0.57	0.57	0.57
Support Vector Classifier (SVC)	0.62	0.63	0.62	0.61

Table 2. Confusion Matrix for All Experiments

Experiment	Negative	Positive	Neutral
Random Forest (RF)			
Negative	101	32	98
Positive	25	204	83
Neutral	43	39	215
Logistic Regression (LR)			
Negative	125	44	62
Positive	44	218	50
Neutral	67	49	181
K-nearest Neighbors (KNN)			
Negative	92	68	71
Positive	61	150	101
Neutral	66	65	166
Bernoulli Naïve Bayes (BNB)			
Negative	81	31	119
Positive	29	215	68
Neutral	36	37	224
Gradient Boosting (GB)			
Negative	62	31	138
Positive	16	179	117
Neutral	27	49	221
Decision Tree (DT)			
Negative	105	38	88
Positive	44	202	66
Neutral	58	65	174
Support Vector Classifier (SVC)			
Negative	87	46	98
Positive	22	231	59
Neutral	27	59	211

5. Discussion and Conclusion

In this study, we conducted a series of experiments using different ML algorithms and text features for sentiment analysis. Table 1 represents a summary of our experiments and shows that Logistic Regression (LR) and Support Vector Classifier (SVC) have the highest F1 score with 0.62 followed by Random Forest (RF) and Bernoulli Naïve Bayes (BNB) with 0.61 F1 score. Table3 represents Accuracy, Precision, Recall, and F1 score performance metrics for each class separately; Positive class, Negative class, and Neutral class for the algorithms conducted in the study. Table3 shows that the best achieved F1 score result is for BNB algorithm for the positive class with 0.72 followed by SVC with 0.71 for the positive mood as well. Positive class's lowest F1 score is for KNN with 0.50 followed by GB with 0.63. For the Negative class, the best achieved F1 score was for LR with 0.54 followed by RF with 0.51, and the lowest performance was for GB with 0.37 F1 score followed by KNN with 0.41. For the Neutral class, the highest F1 score was for SVC and BNB with 0.64 followed by RF with 0.62, and the lowest F1 score was for KNN with 0.52 followed by DT with 0.56. SVC, BNB, and DT worked better with the Positive mood tweets, while LR worked better with the Negative mood tweets, BNB, GB, and SVC worked better for the Neutral mood tweets.

To discuss the experiments in detail, we firstly used the RF algorithm; this algorithm classifies text data by combining many decision trees after converting them into vector space. From Table 3, we observed that the RF model has one of the lowest accuracies in the study for the Negative class with 0.34. However, it achieved one of the highest accuracies as show in Table1 and achieved a moderate performance for the Positive class with 0.53 and Neutral class with 0.45 as shown in Table3. This can be attributed to the class imbalance in the data set and the differences in the characteristics of different classes. For example, in the confusion matrix in Table2, the RF model, Positive class (204) appears to have a moderate level compared with other algorithms. However, we can see that there are also incorrect predictions for Negative class and Neutral class as well.

Table 3. Performance Metrics for Negative, Positive, and Neutral Classes

Experiment	Accuracy	Precision	Recall	F1 score
Random Forest (RF)				
Negative Class	0.34	0.60	0.44	0.51
Positive Class	0.53	0.74	0.65	0.70
Neutral Class	0.45	0.54	0.72	0.62
Logistic Regression (LR)				
Negative Class	0.54	0.53	0.54	0.54
Positive Class	0.70	0.70	0.70	0.70
Neutral Class	0.61	0.62	0.61	0.61
K-nearest Neighbors (KNN)				
Negative Class	0.40	0.42	0.40	0.41
Positive Class	0.48	0.53	0.48	0.50
Neutral Class	0.56	0.49	0.56	0.52
Bernoulli Naïve Bayes (BNB)				
Negative Class	0.35	0.56	0.35	0.43
Positive Class	0.69	0.76	0.69	0.72
Neutral Class	0.75	0.55	0.75	0.64
Gradient Boosting (GB)				
Negative Class	0.27	0.59	0.27	0.37
Positive Class	0.57	0.69	0.57	0.63
Neutral Class	0.74	0.46	0.74	0.57
Decision Tree (DT)				
Negative Class	0.45	0.51	0.45	0.48
Positive Class	0.65	0.66	0.65	0.66
Neutral Class	0.59	0.53	0.59	0.56
Support Vector Classifier (SVC)				
Negative Class	0.38	0.64	0.38	0.47
Positive Class	0.74	0.69	0.74	0.71
Neutral Class	0.71	0.57	0.71	0.64

Next, we used the LR algorithm. This algorithm classifies text data with a linear model. The best achieved accuracy among all experiments goes for this algorithm with 0.62 shared with SVC (Table1). From the detailed analysis for each class shown in Table3, we observed that the LR model has the second highest accuracy for the Positive class with 0.70, and the second highest accuracy also for the negative class with 0.54, and the fourth highest accuracy for the Neutral class with 0.61. From the statistics, we can see that this algorithm works better with classes that carry emotions; this may be due to the class distribution and the fact that the model has a linear classification capability.

We also tried the KNN algorithm. This algorithm classifies a new data point based on the majority of its closest neighbours. In our experimental results, we observed that the KNN model has lower accuracy values for all classes compared with the first other algorithms. The accuracy value for the positive class is not only considered to be the lowest with 0.50, but it is also far away from the next one, which is 0.63. This may be due to the fact that the KNN algorithm is sensitive to class balance and similarities of text data.

We continued in evaluated the sentiment analysis task using BNB, GB, DT, and SVM machine learning algorithms in our experiments. We observed that each algorithm yielded different performance results for different classes as shown in Table3. Algorithms such as RF, SVC, RF, and BNB, all achieved reasonable accuracy rates in the test data set. In particular, the LR models have emerged as an effective option to achieve a high level of success in sentiment analysis. KNN, on the other hand, was the model with the lowest accuracy rate. Some algorithms showed a sensitivity to certain classes compared with the other two classes; SVC showed a sensitivity toward Negative class of 0.38 compared with 0.74 and 0.71 for the positive and neutral classes, respectively and BG showed the same sensitivity with 0.27 for the negative class vs. 0.57 and 0.74 for the positive and neutral classes. However, accuracy alone may not be enough to fully understand the performance of a model. In some cases, factors such as class imbalance must be considered. That's why it's important to evaluate along with other metrics. For example, if the classes are unbalanced and the majority of correctly classified samples belong to the majority class, the accuracy rate can be misleading. In this case, other metrics such as precision, recall, or F1 score must also be taken into account.

Confusion matrix analysis (Table3) for KNN shows that 0.48 of Positive moods and 0.40 of Negative moods were correctly classified. The GB algorithm, on the other hand, shows that 0.57 of Positive moods and 0.27 of Negative moods were correctly classified. As a result, it was found that deep learning-based algorithms such as LR and RF were the most effective options in sentiment analysis in Turkish texts. However, traditional algorithms such as SVC also has a reasonable accuracy rate and may be preferable, especially for fast classification. KNN and GB, on the other hand, are alternatives that can be used in some cases and can be used for a comparison purposes. Our findings make an important contribution to future studies aimed at identifying the most appropriate algorithms for sentiment analysis in Turkish texts and to better understand emotional content. Confusion matrix analysis also helped us to evaluate the emotion classification performance of the algorithms in more detail for each mood.

Another important aspect that might be taken in consideration in understanding the result properly is the characteristics of the data set, and the nature of the algorithms; further work and optimization can be done to improve performance. The results we obtained in the sentiment analysis were quite satisfactory. In the classification task performed on the different machine learning algorithms used in the analysis, it was observed that all models performed at acceptable levels except KNN.

As a result, the accuracy rate alone allows us to make an assessment based on a model, but other factors and metrics must also be considered. However, there are also some challenges encountered in the sentiment analysis process. For example, some data points may show ambiguity due to multiple interpretations of certain emotional expressions. These uncertainties can create difficulty in accurate classification and increase the likelihood of errors in results. For example, situations where positive emotional expressions are more common than negative expressions. This can make it difficult to accurately classify the model's bias due to imbalance and underrepresented classes. In addition, language features such as variability in language, idioms, puns, and irony can also affect correct classification.

It is important to experiment with feature engineering strategies to increase accuracy. In addition, the accuracy rate should be evaluated depending on the requirements of our data set and the purpose of our analysis. When similar literature related to the field of Sentiment Analysis is examined, it is seen that the first study was conducted by Pang et al. [16], and movie reviews in the Internet Movie Database archive were used as a data set. In their study, they created the vector space models required for classification by using feature extraction methods such as unigram, bigram, Part of Speech (POS) on the relevant data set. Tokcaer [2] performed the classification process using machine learning algorithms such as Naive Bayes, Maximum Entropy and SVM on the data set obtained as a result of vector space models. As a result of the findings, the best result in the classification of sentiment analysis was obtained by the SVM machine learning method with an accuracy rate of 82.9% on the unigram data set. In addition, O'Connor et al. [17] applied sentiment analysis to comments on twitter and health-related forums to investigate patients' negative thoughts about the side effects of medications. In practice, the machine learning-based ADRMine method, which uses conditional random fields, was used to extract the concepts in the field of medicine, which were also put forward by them. 6279 and 1784 comments from the health site DailyStrength and Twitter were used as a data set. When the results were examined, it was observed that the ADRMine method gave a higher success rate than SVM and MetaMap methods, which are classifiers used in the field of health, with 82.1%.

As a conclusion, a 62% accuracy rate is not a sufficient metric to evaluate the success of an analysis. Further studies can be done to better understand and improve the analysis results. To improve any model, it's important to experiment with different algorithms, further review the data set, and evaluate other performance metrics. A 62% accuracy rate can be a starting point, but it's important to evaluate other factors to better understand and refine your analysis and model. A 62% accuracy rate is an acceptable result based on the purpose of our analysis and the context.

As a result, our experiments have shown that machine learning models can be used effectively in the field of sentiment analysis. These models provide a valuable tool for classifying text data and understanding emotional content. However, it is important to consider challenges such as ambiguities and language characteristics. In the future, we aim to obtain more precise results with experiments and model improvements with larger and more diverse data sets. Negation is one of the most important concepts that affects the accuracy of the model, in this study "I liked" and "I don't like" are both classified as positive sentiment while after using negation the second sentence "I don't like" should be classified part of the negative class. Another enhancement that is suggested for future work is to use lemmatization instead of stemming in the text preprocessing step, stemming sometimes is harsh and affect the sentiment of the word, so using lemmatization may lead to enhancement in the model performance.

For future work also, studying SA in Turkish data sets from Twitter and compare it with data sets from other domain will be interesting for the reason that writing reviews for hotels, hospitals, restaurants, etc is different than writing tweets. Tweets are normally shorter and classifying them has its own challenge, while reviews are more detailed and contains direct content related to certain good/service and the goal behind writing the review is either to give the opinion or might be used as a complain. Alawi and Bozkurt [18] and Cam et al. [19] focused on data sets from Twitter, while Inan [20] and Alzoubi et al. [21] focused on data sets from reviews. In Alawi and Bozkurt [18], the conventional machine learning model SVM achieved an accuracy of 0.8805 and an F1-Score of 0.8348, and in Cam et al. [19], SVM and Multilayer Perceptron classifier achieved 0.89 and 0.88 accuracy rates. In Inan [20], the logistic regression method was the most successful classification algorithm, with an accuracy rate of 0.92, and in Alzoubi et al. [21] the best achieved accuracy in traditional techniques was 78% accuracy for the Support Vector Machine. Comparing SA classification techniques for data sets collected from Twitter with other data sets resources might give some insights for research in the domain.

References

[1] Tokcaer, S. (2021). Türkçe metinlerde duygu analizi. Yaşar University E-Dergisi, 16(63), 1514-1534.

[2] Kaynar, O., Görmez, Y., Yıldız, M., & Albayrak, A. (2016, September). Makine öğrenmesi yöntemleri ile Duygu Analizi. In International Artificial Intelligence and Data Processing Symposium (IDAP'16) (Vol. 17, No. 18, pp. 17-18).

[3] Danisman, T., & Alpkocak, A. (2008, April). Feeler: Emotion classification of text using vector space model. In AISB 2008 convention communication, interaction and social intelligence (Vol. 1, p. 53). T.

[4] Medler, D. A., Arnoldussen, A., Binder, J.R., & Seidenberg, M.S. (2005). The Wisconsin Perceptual Attribute Ratings Database. http://www.neuro.mcw.edu/ratings/

[5] Alpkoçak, A., Tocoglu, M. A., Çelikten, A., & Aygün, İ. (2019). Türkçe metinlerde duygu analizi için farklı makine öğrenmesi yöntemlerinin karşılaştırılması. Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, 21(63), 719-725.

[6] Türkmenoğlu, C. (2016). Türkçe metinlerde duygu analizi (Doctoral dissertation, Fen Bilimleri Enstitüsü).

[7] Kozareva, Z., Navarro, B., Vázquez, S., & Montoyo, A. (2007, June). UA-ZBSA: a headline emotion classification through web information. In Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007) (pp. 334-337).

[8] Mohammad, S. (2012, June). Portable features for classifying emotional text. In Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 587-591).

[9] Chaffar, S., & Inkpen, D. (2011). Using a heterogeneous dataset for emotion analysis in text. In Advances in Artificial Intelligence: 24th Canadian Conference on Artificial Intelligence, Canadian AI 2011, St. John's, Canada, May 25-27, 2011. Proceedings 24 (pp. 62-67). Springer Berlin Heidelberg.

[10] Boynukalın, Z. (2012). Emotion analysis of Turkish texts by using machine learning methods (Master's thesis, Middle East Technical University).

[11] Kaya, M., Fidan, G., & Toroslu, I. H. (2012, December). Sentiment analysis of Turkish political news. In 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (Vol. 1, pp. 174-180). IEEE.

[12] Oflazer, K. (1994). Two-level description of Turkish morphology. Literary and linguistic computing, 9(2), 137-148.

[13] Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2007, January). Analysis of affect expressed through the evolving language of online communication. In Proceedings of the 12th international conference on Intelligent user interfaces (pp. 278-281).

[14] Wang, W., Chen, L., Thirunarayan, K., & Sheth, A. P. (2012, September). Harnessing twitter" big data" for automatic emotion identification. In 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing (pp. 587-592). IEEE.

[15] U. Erogul. Sentiment analysis in Turkish. Master's thesis, Middle East Technical University, 2009.

[16] Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. arXiv preprint cs/0205070.

[17] O'Connor, K., Pimpalkhute, P., Nikfarjam, A., Ginn, R., Smith, K. L., & Gonzalez, G. (2014). Pharmacovigilance on twitter? Mining tweets for adverse drug reactions. In AMIA annual symposium proceedings (Vol. 2014, p. 924). American Medical Informatics Association.

[18] Alawi, A. B., & Bozkurt, F. (2024). A hybrid machine learning model for sentiment analysis and satisfaction assessment with Turkish universities using Twitter data. Decision Analytics Journal, 11, 100473.

[19] Cam, H., Cam, A. V., Demirel, U., & Ahmed, S. (2024). Sentiment analysis of financial Twitter posts on Twitter with the machine learning classifiers. Heliyon, 10(1).

[20] İnan, H. E. Comparison of Machine Learning Algorithms for Classification of Hotel Reviews: Sentiment Analysis of TripAdvisor Reviews. GSI Journals Serie A: Advancements in Tourism Recreation and Sports Sciences, 7(1), 111-122.

[21] Alzoubi, Y. I., Topcu, A. E., & Erkaya, A. E. (2023). Machine learning-based text classification comparison: Turkish language context. Applied Sciences, 13(16), 9428.

■ RESEARCH ARTICLE

Deep Learning Models for the Detection and Classification of COVID-19 and Associated Lung Diseases Using X-Ray Images

Osman Dikmen [a] †

[a] Department of Electric-Electronics, Düzce University, Düzce, Türkiye
† osmandikmen@duzce.edu.tr

RECEIVED SEPTEMBER 17, 2024
ACCEPTED SEPTEMBER 24, 2024

CITATION Dikmen. O. (2024). Deep learning models for the detection and classification of COVID-19 and associated lung diseases using X-ray images. *Artificial Intelligence Theory and Applications, 4(2*), 121-142.

Abstract

The COVID-19 pandemic has introduced exceptional challenges to healthcare systems worldwide, underscoring the urgent need for swift and precise diagnostic solutions. In this research, we investigate the performance of various deep learning models, including VGG19, ResNet18, and a ResNet18-based U-Net, as well as a custom Convolutional Neural Network (CNN) developed in MATLAB, for the classification and segmentation of lung X-ray images. The dataset includes X-ray images from individuals diagnosed with COVID-19, viral pneumonia, lung opacity, and healthy individuals. The dataset was divided into 80% for training and 20% for testing, with data augmentation techniques implemented to enhance the model's effectiveness. The VGG19 model, utilizing transfer learning, demonstrated strong diagnostic capabilities, achieving high accuracy rates for COVID-19, lung opacity, healthy lungs, and viral pneumonia classification, with a test accuracy of 97.5%. ResNet18 was employed for both classification and as part of a hybrid model incorporating a U-Net-inspired decoder for lung disease segmentation. The ResNet18 model achieved competitive accuracy and loss metrics, while the ResNet18-based U-Net model excelled in image segmentation tasks, demonstrating its potential in biomedical image analysis. Additionally, a customized CNN model was developed using MATLAB for the classification of the four lung conditions. This model produced visual outputs, including training-validation loss/accuracy graphs and confusion matrices. Our results indicate that deep learning models, especially when combined with transfer learning and customized architectures, offer a powerful approach to diagnosing COVID-19 and related lung conditions. Future work will focus on refining these models with larger datasets and further experimentation to enhance diagnostic performance across diverse clinical settings.

Keywords: COVID-19, deep learning, medical image classification, x-ray imaging, convolutional neural networks

1. Introduction

The COVID-19 pandemic emerged as a result of the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) virus, which triggered a profound global health emergency. Initially identified in the city of Wuhan in late 2019, the disease swiftly

Artificial Intelligence Theory and Applications, ISSN: 2757-9778. ISBN: 978-605-69730-2-4 © 2024 İzmir Bakırçay University

evolved into a global health crisis. COVID-19 presents a diverse array of clinical symptoms, ranging from mild to severe respiratory distress. Swift and precise diagnosis for this illness is essential for controlling its spread and determining appropriate treatment strategies [1].

The virus rapidly spread from Wuhan, reaching numerous countries around the globe. Its impact was notably felt in major regions, including North America, parts of South Asia, South America, Western Europe, and Eastern Europe In March 2020, the World Health Organization (WHO) in a formal manner announced COVID-19 as a pandemic, designating it as a worldwide health emergency [1]. During this time, billions of people were required to remain indoors, and many countries enforced lockdowns. By May 19, 2022, there had been approximately 525,080,438 confirmed cases of COVID-19 across more than 219 countries, with 484,920,117 recoveries and 6,294,856 fatalities reported [1].

Diagnosing COVID-19 usually involves several methods: Serology (Antibody) for detecting antibodies, Genetic Real-Time Reverse Transcription Polymerase Chain Reaction (RT-PCR) for genetic material analysis, and Antigen testing [2]. RT-PCR is widely regarded as the gold standard for detecting the coronavirus, delivering reliable results, especially during the initial phases of the infection [3], [4]. Nonetheless, this method has several drawbacks, including issues with false sampling, accessibility, specificity, cost, and extended turnaround times [5]. Additionally, many countries struggle with providing sufficient RT-PCR test kits.

The antibody test detects the presence of IgG and/or IgM antibodies through samples obtained from blood, serum, or plasma [6]. This test is designed to reveal whether an individual has been previously infected and how their immune system responded, but it cannot confirm a current infection. Antibodies typically begin to form between one to three weeks following the onset of symptoms [7]. This method is also costly and requires significant time.

The antigen test attempts to detect the presence of coronavirus infection by collecting samples from nasal swabs. It is less time-consuming and relatively inexpensive compared to other tests [8]. Given these limitations, there is a growing need to explore different diagnostic methods which are both accurate and accessible.

Medical imaging, including X-rays and Computed Tomography (CT), has proven essential in the rapid detection of lung abnormalities associated with COVID-19. X-rays, in particular, are widely accessible and commonly used in clinical settings. Recent advancements in deep learning have further enhanced the utility of medical imaging for disease diagnosis. Convolutional Neural Networks (CNNs), an essential element of deep learning, have shown remarkable accuracy in analyzing medical images. However, it may miss active infections when compared to RT-PCR tests.

Although RT-PCR testing is crucial among traditional diagnostic approaches, the demand for quick and precise alternative methods is growing due to the lengthy duration required for these tests, the need for laboratory facilities, and occasional false-negative results. In this regard, medical imaging techniques become essential. Imaging techniques, including X-rays, are valued for their swift accessibility and extensive application in clinical practice. Similarly, Computed Tomography (CT) imaging is preferred for its expedited and comprehensive diagnostic capabilities.

In this study, we explore several deep learning models for the classification and segmentation of lung diseases, including COVID-19. The VGG19 architecture, renowned for its deep and wide network, is employed through transfer learning to classify X-ray images of COVID-19, lung opacity, viral pneumonia, and healthy lungs. Additionally, a ResNet18-based encoder is combined with a U-Net-inspired decoder to leverage both effective feature extraction and high-resolution segmentation for lung disease analysis. Furthermore, a customized Convolutional Neural Network (CNN) model was developed in MATLAB to classify the four lung conditions, offering a tailored approach to disease detection

Through these models, we aim to address the growing need for quick, precise, and accessible diagnostic tools for COVID-19 and related lung diseases. Our work demonstrates the capability of deep learning techniques in revolutionizing medical diagnostics, offering both high accuracy and clinical applicability in a time of global health crisis.

2. Related Studies

Various deep learning architectures have been introduced to diagnose COVID-19. In Nayak et al. [9], the researchers presented a compact CNN approach called LW-CORONet. This method incorporates convolutional layers, pooling layers, two fully connected (FC) layers, and a rectified linear unit (ReLU) activation function. With its five learnable layers, this architecture aids in extracting crucial features from CXR images.

Gupta and Bajaj [10] created a robust model for automatically detecting COVID-19 by utilizing deep learning (DL) techniques and chest CT scans. They incorporated two pre-trained DL models, DarkNet19 and MobileNetV2, alongside a lightweight DL approach, using publicly available CT scan visual data for concerning automated recognition of COVID-19.

A novel technique has been proposed to advance the classification and screening of COVID-19 patients through chest X-ray (CXR) imaging. This approach integrates cutting-edge deep learning models with refined image analysis methods, aiming to significantly boost diagnostic precision and speed. The method described combines standard data augmentation strategies with generative adversarial networks (GANs) to address data limitations. Additionally, it incorporates various filter banks, including Gabor filters, Sobel filters, and the Laplacian of Gaussian (LoG), to achieve more comprehensive feature extraction [11].

Basu et al. [12] showcased a deep learning approach designed to ascertain the presence of COVID-19 through the assessment of chest X-ray images. This solution utilizes a model pre-trained on a small chest X-ray dataset and domain extension transfer learning. Specifically focusing on identifying cases of COVID-19, this method identifies the specific segments analyzed in classification using Gradient Class Activation Map, thus ensuring transparency in the detection process. The authors reported an overall accuracy rate of

90.1%. However, the proposed system can detect COVID-19 based on a restricted dataset.

Conversely, the ResNet-101 CNN model utilized by Azemin et al. [13] is notable within deep learning strategies. This approach utilized a vast number of images in the pre-training phase to extract crucial features, subsequently re-training the model to identify anomalies in chest X-ray images. Despite these efforts, the method achieved a reported accuracy rate of only 71.9%.

In another study, the MobileNetV2 deep learning model and k-nearest neighbor (k-NN) algorithm were used to detect brain tumors from MRI images, achieving an accuracy rate of 96.44% [14].

For lung cancer detection using CT images, a Convolutional Neural Network (CNN) model was outlined, demonstrating superior performance, achieving a high accuracy rate and fewer layers compared to existing deep learning models [15].

To tackle two distinct multi-class classification challenges, the Xception transfer learning method was employed. The first challenge involved distinguishing among control cases, COVID-19, and various forms of pneumonia, including viral and bacterial types. The second challenge focused on differentiating between control cases, COVID-19, and pneumonia overall. An undersampling technique was implemented to mitigate dataset imbalance by randomly removing samples from the more numerous classes. The dataset comprised 290 chest X-ray images for COVID-19, 310 for control cases, 330 for bacterial pneumonia, and 327 for viral pneumonia. The study reported an accuracy of 89% for classifying all four conditions and 94% for distinguishing between the three categories. For the dataset focusing on three categories, the accuracy was noted as 90% [16].

X-ray imaging is commonly employed to investigate conditions such as fractures, bone misalignments, pneumonia, and tumors. This method has a long history of use, providing a rapid assessment of the lungs and proving beneficial for identifying infections, including COVID-19 [17], [18].

X-rays can produce images showing lung damage, such as pneumonia brought about by the SARS-CoV-2 virus [19]. Due to their speed and low cost, X-rays can help prioritize patients in areas where the healthcare system is strained or access to complex technologies is limited. Furthermore, portable X-ray devices, which are easily transportable with ease to the needed location, are available [19].

CT scans, on the other hand, use X-ray principles to examine soft tissues in the body and are ideal for providing detailed images of organs and soft tissues [20]. Moreover, as they use less radiation, X-rays are faster, less harmful, and more economical than CT scans [21].

Narin et al. [17] introduced a method for the automatic uncovering of COVID-19 through the use of chest X-rays and convolutional neural networks (CNNs). Similarly, Apostolopoulos et al. [18] developed an automatic detection system for COVID-19, which involved analyzing and classifying three different categories: COVID-19, typical pneumonia, and normal conditions.

Transfer learning involves utilizing models that have been previously trained on one task to enhance performance on a new, related task on large and diverse datasets are re-trained on smaller and specific datasets. Using this technique, the model is able to learn faster and more accurately, particularly in fields with limited datasets like medical imaging. The VGG19 model holds significant importance in deep learning research. With its deep and wide layer architecture, VGG19 can classify complex images successfully. Therefore, in this study, the VGG19 model was used to identify a total of 2117 COVID-19 and related lung images.

3. Model Architecture and Methodology

This section outlines the methods and design of three different models used to classify COVID-19 and other lung diseases: the VGG19-based classifier model, the ResNet18-based encoder with a U-Net inspired decoder, and the MATLAB-based customized CNN model. Each model has been developed and evaluated on the dataset that consists of chest X-ray images across four categories: COVID-19, Normal, Lung Opacity, and Viral Pneumonia.

3.1. VGG19-Based Classifier Model

3.1.1 Model Architecture and Transfer Learning

The VGG19 model was employed for classifying the four classes (COVID-19, Normal, Lung Opacity, Viral Pneumonia). It leveraged transfer learning, with pre-trained weights on the ImageNet dataset. For classification purposes, the final layers of VGG19 were fine-tuned to accommodate four output classes. This enabled the model to adapt quickly by utilizing the generalized features learned from large-scale datasets.

3.1.2 Dataset and Preprocessing

The dataset utilized in this study includes chest X-ray images from COVID-19 patients, those with viral pneumonia, and healthy individuals, sourced from multiple databases. These images are available at:
https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database.

A series of preprocessing steps were undertaken to process and prepare the data for the training phase. The dataset consists of 21,165 chest X-ray images, which were classified into four categories:

- COVID-19: 3,616 images
- Normal: 10,192 images
- Lung Opacity: 6,012 images
- Viral Pneumonia: 1,345 images

Table 1. Organization of the Data Set.

Class	Training Set	Test Set
COVID-19	2893	723
Viral Pneumonia	1076	269
Healthy	8154	2038
Lung Opacity	4810	1202
Total	16933	4232

As shown in Figure 1, the dataset captures a wide variety of cases across different lung conditions. The dataset was split into 80% for training (16,932 images) and 20% for testing (4,233 images). All images were resized to 224x224 pixels to meet the input size requirements of the VGG19 model. Pixel values were normalized to the range of 0-1 for consistency in model processing.

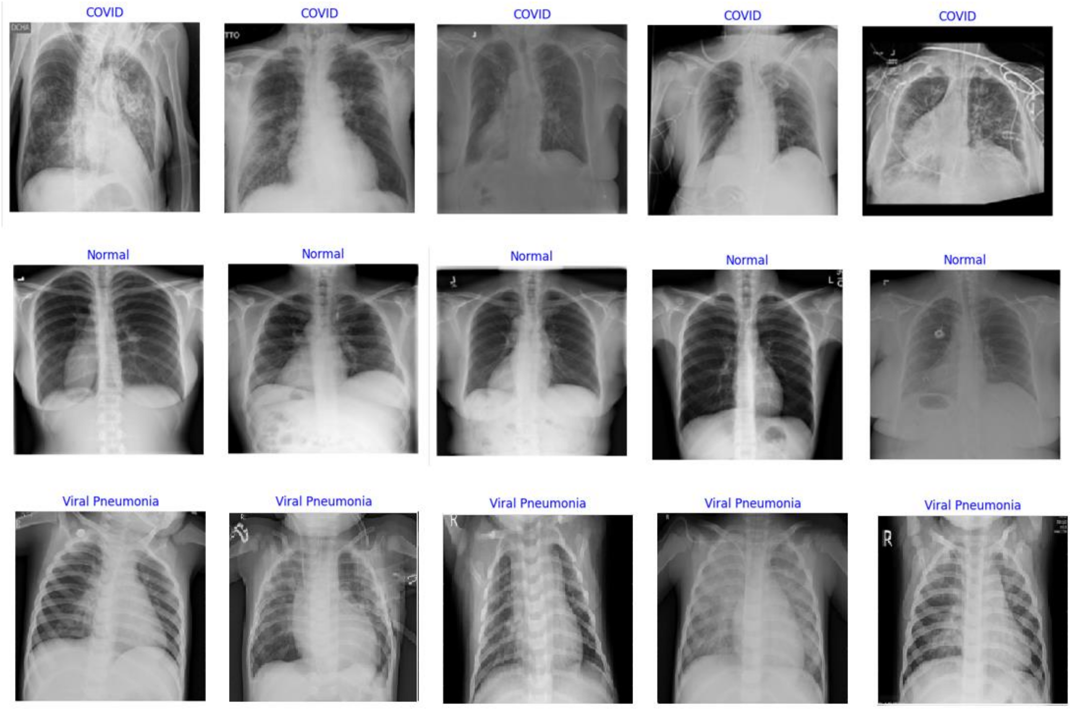


Figure 1. Example images utilized in the experimental analysis of this study.

3.1.3 Data Augmentation

To reduce overfitting and improve the model's ability to generalize, several data augmentation strategies were employed: Horizontal flipping, Random rotation, Zooming and shifting, Brightness adjustments. These augmentations allowed the model to become more invariant to orientation and lighting changes, improving its real-world applicability.

3.1.4 Training Process

The VGG19 model was optimized with the Adam algorithm, configured with a learning rate of 0.0001 to enhance its training efficiency. The training process was monitored with early stopping to prevent overfitting, and the model with the best validation accuracy was saved. Transfer learning enabled faster convergence by leveraging pre-trained weights. An evaluation was conducted on the model's performance using accuracy, precision, specificity, and F1 score metrics. These metrics were calculated using the following equations:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{4}$$

Where TP is true positives, TN is true negatives, FP is false positives and FN is false negatives.

3.2. ResNet18-Based Encoder and U-Net Inspired Decoder Model

3.2.1 Encoder Architecture (ResNet18)

The encoder component of this model is based on ResNet18, a residual network that uses residual connections to facilitate learning in deep networks. The network begins with a 7x7 convolutional layer, followed by several 3x3 convolutional blocks and max-pooling layers. Residual connections ensure the gradient flows effectively through the network, improving convergence. To mitigate the learning difficulties in deep networks, residual connections are utilized. The residual block is defined as

$$\mathrm{Y} = \mathrm{F}(\mathrm{X}) + \mathrm{X} \tag{5}$$

where X is the input and F is the learned transformation.

To reduce feature dimensions, max pooling is applied:

$$\mathrm{X}_{\mathrm{pool}}=\mathrm{MaxPool}(\mathrm{X},\text{kernel size}=2) \tag{6}$$

3.2.2 Decoder Architecture (U-Net)

The decoder is inspired by the U-Net architecture, which is well-suited for segmentation tasks. The decoder uses transpose convolutional layers to upsample the feature maps to their original dimensions. These upsampled features are further refined by dual convolutional layers, and the final output is generated through a convolutional layer that produces a segmentation map of the lung regions, essential for differentiating between COVID-19 and other lung diseases. These layers upsample the feature maps to the original dimensions. Mathematically, this operation is represented as;

$$\mathrm{X}_{\mathrm{up}}=\mathrm{ConvTranspose2d}(\mathrm{X}_{\mathrm{in}},\mathrm{W}_{\mathrm{trans}},\text{stride}=2) \tag{7}$$

These layers refine the upsampled features for detailed segmentation:

$$\mathrm{X}_{\mathrm{out}}=\mathrm{Conv2d}(\mathrm{X}_{\mathrm{up}},\mathrm{W}_{\mathrm{conv}},\text{kernel size}=3) \tag{8}$$

This produces the final segmentation map:

$$Y_{seg}=\text{Conv2d}(X_{act},W_{conv},\text{kernel size}=1) \quad (9)$$

3.2.3 Training Process

The ResNet18 encoder and U-Net decoder model was trained to perform lung image segmentation. The model was optimized using the Adam optimizer, with careful tuning of the learning rate. The model was evaluated on its ability to identify lung regions affected by COVID-19 and other lung conditions, improving the overall diagnostic accuracy for the dataset. The loss function used is Binary Cross Entropy with Logits, defined as [22]:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log \hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i)] \quad (10)$$

where $\hat{y}_i$ represents the model's prediction, y_i denotes the true label, and N represents the total number of samples.

The model is trained using the Adam optimization algorithm, which updates weights according to:

$$W_{t+1} = W_t - \eta.\nabla Loss(W_t) \quad (11)$$

where η is the learning rate, W_t denotes the model weights, and $.\nabla Loss(W_t)$ represents the gradient of the loss function.

3.3. Matlab-Based Customized CNN Model

3.3.1 Model Architecture

A customized CNN model was designed using Matlab for the classification task. The input to the model consists of RGB images of size 256x256x3, and the architecture includes:

Input Layer: The model accepts RGB images of size 256x256x3.

Convolutional Layers: The initial convolutional layer has 32 filters with a 3x3 kernel size. This layer is succeeded by batch normalization and ReLU activation layers, which are employed to improve the effectiveness of feature extraction.

$$Output_i=ReLU(Conv(Input,Filters_i,Stride_i)+Bias_i \quad (12)$$

Here, Conv denotes the convolution operation, ReLU is the activation function, and Outputi is the output feature map.

Max Pooling Layer: Reduces spatial dimensions to increase computational efficiency. Max pooling is typically performed with 2×22 \times 22×2 filter sizes.

Fully Connected Layers: Enhance the learning capacity of the model and support the extraction of high-level features. The final layer includes a softmax activation function that provides output for four classes.

$$\text{Softmax}(z) = \frac{exp(z_i)}{\sum_j exp(z_j)} \tag{13}$$

Here, z denotes the output of the classification layer and Softmax is the activation function.

3.3.2 Dataset and Data Augmentation

To prepare the images for the model, several preprocessing techniques were applied. “imageDatastore” was used to split the labeled images into training and testing sets. Data augmentation was performed to enhance the model's generalization capability:

Rotation: Images are rotated at different angles to enable the model to identify and interpret images from various perspectives.

Horizontal and Vertical Flipping: Images are flipped both horizontally and vertically to aid the model in learning and recognizing symmetric patterns.

Zooming and Shifting: Images are zoomed and shifted to allow the model to learn objects at different scales.

3.3.3 Training and Evaluation

The model was trained using the “trainNetwork” function with the “adam” optimization algorithm. Training involved calculating loss and accuracy using the following equations:

Loss Function:

$$\text{Loss} = -\sum_i y_i \log \hat{y}_i \tag{14}$$

Where y_i are the true labels and $\hat{y}_i$ denotes the predicted probabilities.

Accuracy Calculation:

$$\text{Accuracy} = \frac{Correct\ Predictions}{Total\ Examples} \tag{15}$$

Loss and accuracy graphs, as well as confusion matrices, were presented. The loss and accuracy graphs illustrate the learning process of the model, while the confusion matrices detail the classification performance and the rates of correct and incorrect classifications for each class.

4. Experimental Results

The experimental results for the three models—VGG19-based classifier, ResNet18-based encoder with U-Net inspired decoder, and the Customized Convolutional Neural Network (CNN)—are presented in this section. These results evaluate the performance of each model across the dataset. VGG19 focusing on key metrics such as accuracy, precision, specificity, F1 score, and loss during both the training and testing phases. Additionally, confusion matrices and visualizations of accuracy and loss curves are presented to demonstrate the models' learning patterns and generalization capabilities.

4.1 VGG19-Based Classifier Model

4.1.1 Performance Metrics

The evaluation of the VGG19 model's performance was conducted using essential metrics such as accuracy, precision, specificity, and the F1 score. Table 2 offers a detailed overview of how the model performed for each class.

Table 2. VGG19 Model Performance Metrics.

Method	Precision (%)	Specificity (%)	F1 Score (%)	Accuracy (%)
COVID-19	98	98	98	97.5
Lung Opacity	95	91	93	91.02
Healthy	94	97	96	97.3
Viral Pneumonia	98	96	97	96.2

The COVID-19 class achieved the highest performance across most metrics, demonstrating the model's effectiveness at distinguishing COVID-19 from other lung conditions.

4.1.2 Accuracy and Loss Curves

The training process for the VGG19 model is depicted in Figure 2 and Figure 3:

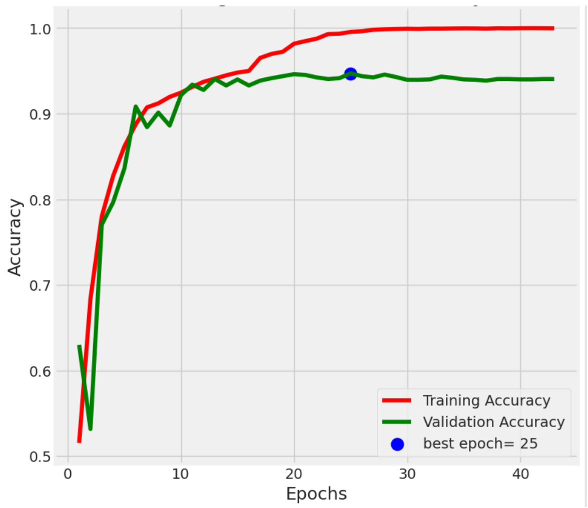


Figure 2. Training and Validation Accuracy Graph.

Figure 2 shows the accuracy progression over the training phase, where the training accuracy consistently increases, while validation accuracy stabilizes after reaching a peak.

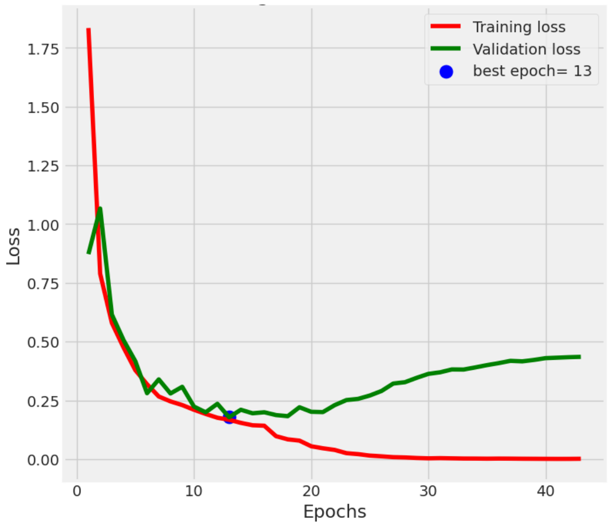


Figure 3. Training and validation loss graph.

Figure 3 illustrates the reduction in loss during training, with the training loss decreasing steadily, and validation loss stabilizing after a certain point. This suggests that the model has learned effectively and is able to generalize well to unseen data.

4.1.3 Confusion matrix

Figure 4 presents the confusion matrix, which provides insights into the classification performance for each class. The matrix shows high counts of true positives (TP) and true negatives (TN) across all classes, indicating strong model performance. The minimal instances of false positives (FP) and false negatives (FN) underscore the model's reliability in accurately distinguishing between COVID-19, lung opacity, healthy individuals, and viral pneumonia cases.

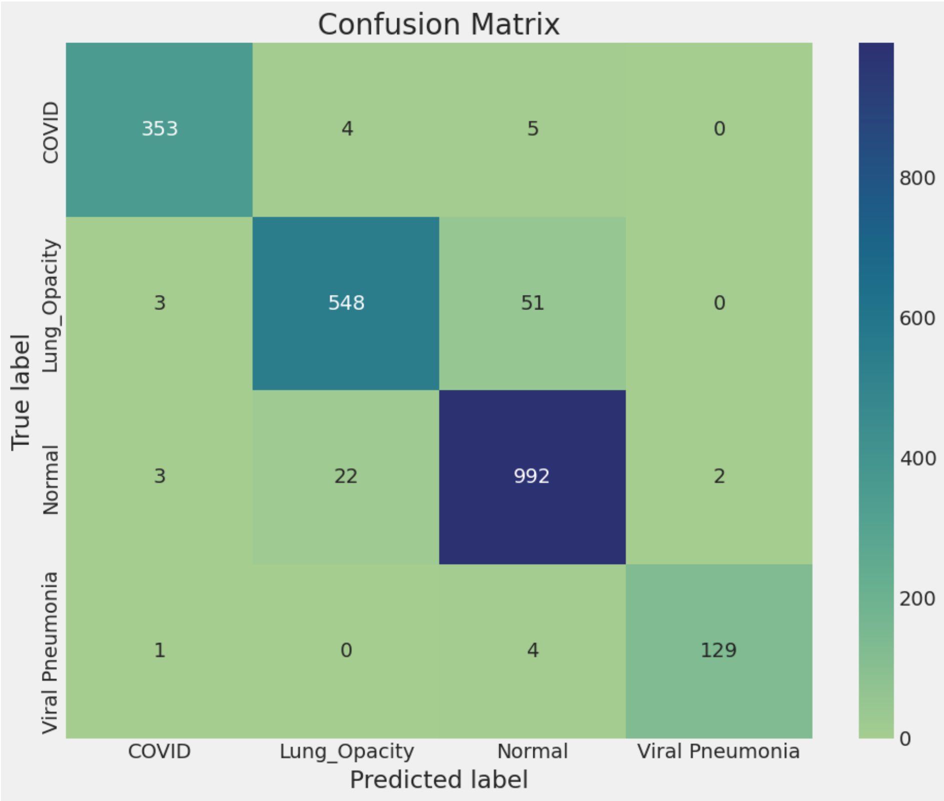


Figure 4. Confusion matrix.

The suggested approach for diagnosing COVID-19 demonstrates superior performance compared to other methods.

Table 4. Evaluating the Proposed COVID-19 Diagnostic Method Against Alternative Approaches.

Study	**Type of Images**	**Number of Cases**	**Method Used**	**Accuracy (%)**
[18]	X-ray	224 COVID-19 700 Pneumonia 504 Normal	Transfer learning+VGG19	93.48
[23]	X-ray	1300 COVID-19 1300 Pneumonia 1300 Normal	DTL+VGG-19	92.92
[24]	X-ray	219 COVID-19 1300 Pneumonia 1300 Normal	BND+VGG-19	95.48
[25]	CT	219 COVID-19 224 Pneumonia 758 Normal	ResNet+Location Attention	86.07
[26]	X-ray / CT	1493 COVID-19 2780 Pneumonia 1538 Normal	Inception Resnet V2	92.18
[27]	X-ray	2210 COVID-19 2340 Pneumonia 1480 Normal	CXRVN	93.07

[28]	X-ray	305 COVID-19 305 Pneumonia 305 Normal	CovXNet	89.6
[29]	CT	777 COVID-19 505 Pneumonia 708 Normal	ARENET	93.00
[30]	X-ray	260 COVID-19 300 Pneumonia 300 Normal	Transfer learning +VGG19	89.30
This Study	**X-ray**	**3616 COVID-19** **1345 Pneumonia** **10192 Normal** **6012 Lung Opacity**	**Transfer learning +VGG19**	**96.00**

4.2 ResNet18-Based Encoder and U-Net Inspired Decoder Model

4.2.1 Performance Evaluation

The ResNet18-based encoder with a U-Net inspired decoder was designed for lung segmentation and feature extraction from the X-ray images. The performance of the model was analyzed using both accuracy and loss metrics during training and validation.

The training and validation loss curves show that the model's training loss decreases consistently, while validation loss remains stable after a certain point, suggesting good generalization. Likewise, the accuracy curves show that training accuracy steadily improves with each epoch, and validation accuracy reaches a plateau, demonstrating that the model is not overfitting.

4.2.2 Segmentation Performance

The ResNet18-U-Net model was specifically evaluated for its capacity to identify and segment lung areas impacted by different conditions such as COVID-19 and viral pneumonia. The model attained strong performance in segmenting lung opacity and viral pneumonia areas, supporting its application in biomedical imaging.

The ResNet18 and ResNet18-Based Encoder with U-Net Inspired Decoder models were evaluated based on their training and validation accuracy, as well as their training and validation loss. The graphs below depict the models' performance during training. The training and validation accuracy graphs illustrate how the models progressively improved their accuracy over time, demonstrating their ability to learn effectively from the data. Likewise, the training and validation loss graphs show a consistent reduction in loss, indicating that both models successfully minimized errors during training while maintaining stable performance on the validation set. These results highlight the models' capability for generalization and effective learning throughout the training process.

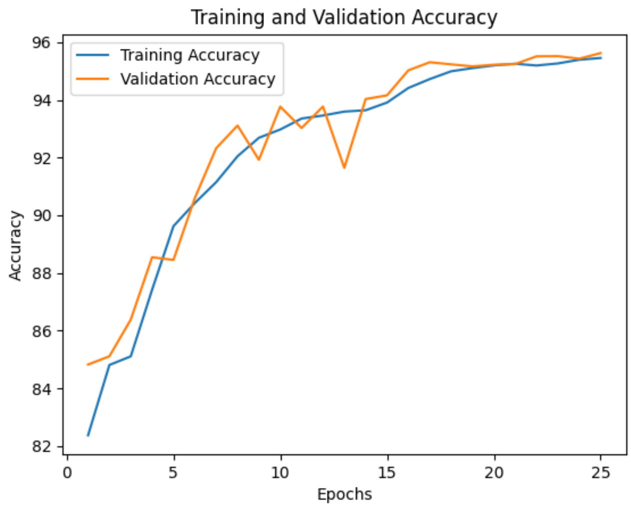


Figure 5. Training and validation accuracy graph for ResNet18.

Figure 6. Training and validation loss graph for ResNet18.

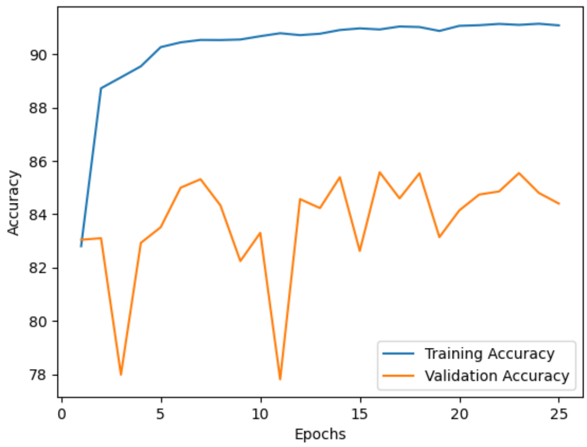


Figure 7. Training and validation accuracy graph for ResNet18-Based Encoder with U-Net Inspired Decoder.

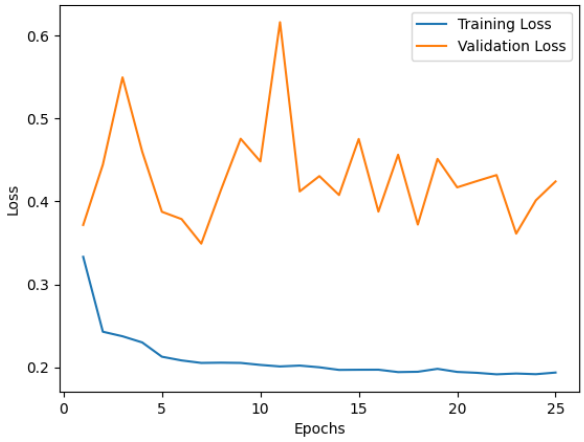


Figure 8. Training and validation loss graph for ResNet18-Based Encoder with U-Net Inspired Decoder.

4.3 Customized Convolutional Neural Network (CNN) Model (Matlab)

4.3.1 Model Performance

The customized CNN model developed in MATLAB was evaluated based on accuracy, with the model achieving an overall accuracy of 85.56% on the test set. The training process involved 2,645 iterations across 5 epochs, with each epoch consisting of 529 iterations. This setup illustrates the model’s ability to accurately differentiate among COVID-19, normal, lung opacity, and viral pneumonia cases.

4.3.2 Accuracy and Loss Graphs

During the training process, the model's loss steadily decreased, indicating successful learning. The accuracy chart illustrates that the model attained high performance on both the training and validation datasets, demonstrating minimal signs of overfitting. The graphs depicting Training and Validation Accuracy and Loss are presented below, offering a visual summary of the model's performance across the training phase.

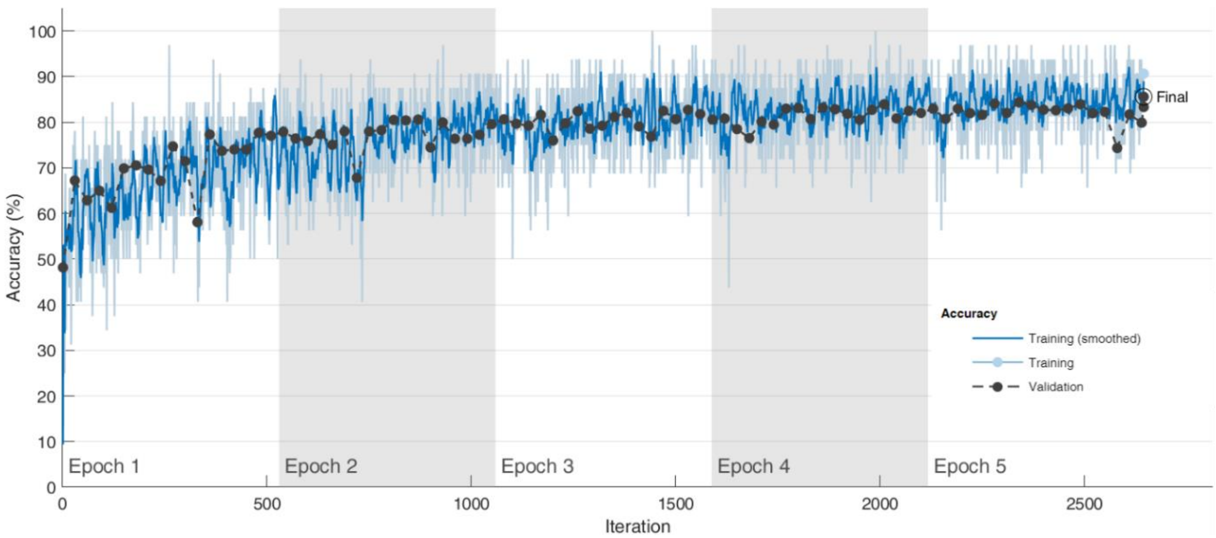


Figure 9. Accuracy vs. Iteration for Matlab.

This figure displays the model's accuracy throughout the training process. The x-axis represents the number of iterations, while the y-axis illustrates the accuracy achieved by the model at each iteration. The graph presents how the model's accuracy evolves as training progresses, showing the improvements made with each iteration and helping to assess the overall effectiveness of the training process. The plot highlights both the training accuracy and the validation accuracy, enabling a comparison of the model's performance on training data versus unseen validation data.

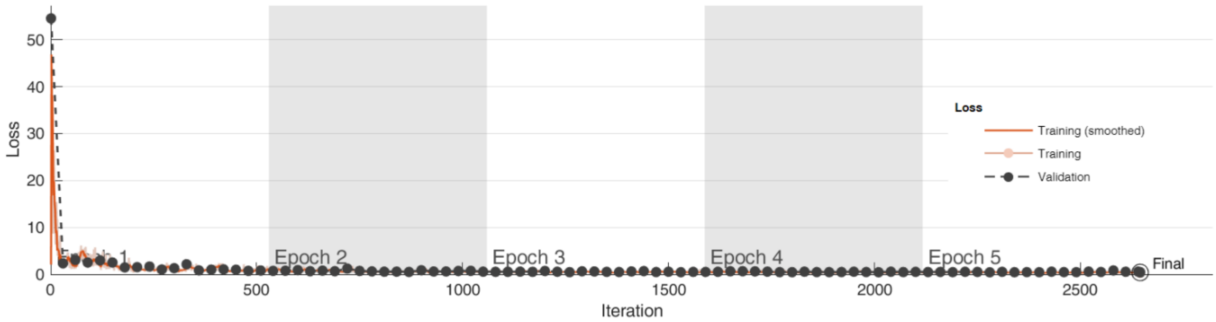


Figure 10. Loss vs. Iteration for Matlab.

This figure shows the model's loss during training. The x-axis signifies the number of iterations, while the y-axis reflects the loss value recorded for the model at each iteration. This plot helps visualize how the loss decreases as the model learns and improves over time. By comparing the loss values across iterations, we can evaluate the convergence of the model and the effectiveness of the training. The graph typically displays both the training loss and the validation loss, providing a clear view of how well the model is generalizing to unseen data as training progresses.

4.3.3 Confusion Matrix

The confusion matrix for the customized CNN model illustrates the performance of the model in classifying the four lung conditions: COVID-19, Normal, Lung Opacity, and Viral Pneumonia. The matrix offers a detailed account of the predictions, categorizing them into true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each class. Each cell in the matrix reflects the number of predictions made by the model for a particular class compared to the actual class.

The diagonal elements represent the correct predictions for each class, showing high precision and recall rates across all classes. This indicates the model's effectiveness in accurately distinguishing between different lung conditions. The off-diagonal elements represent misclassifications, providing insights into the types of errors the model makes. By analyzing these errors, we can assess areas for improvement and understand how well the model generalizes to various conditions.

This detailed confusion matrix is crucial for evaluating the model's overall classification performance and for identifying any biases or specific challenges the model may face in distinguishing between certain classes. The confusion matrix is shown in Figure 11.

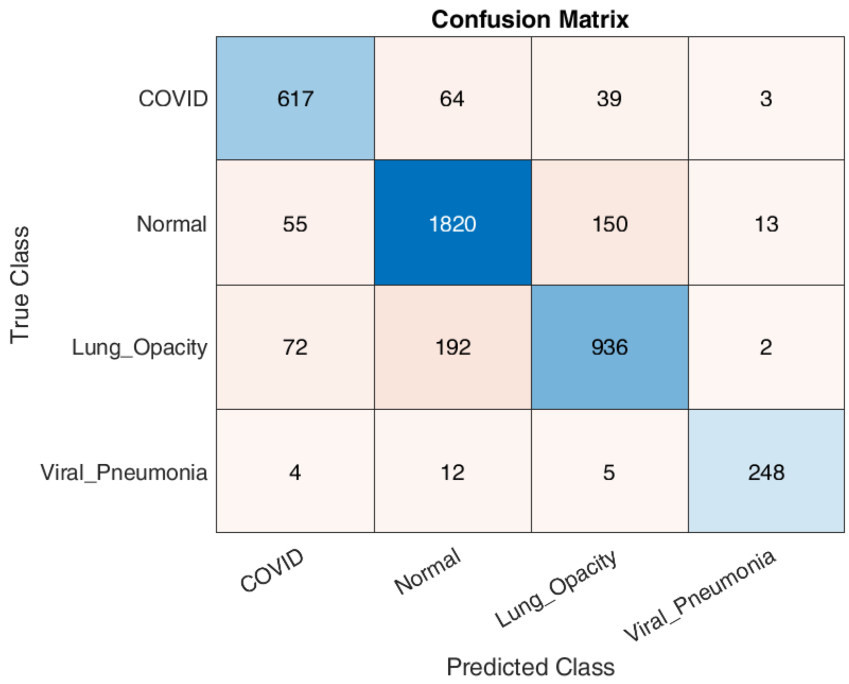


Figure 11. Confusion Matrix for Customized CNN Model (MATLAB).

4.4 Overall Comparison and Discussion

The performance of each model was compared, revealing unique strengths depending on the task:

- The VGG19-based model excelled in classification tasks, particularly in identifying COVID-19 cases with a precision of 98%.
- The ResNet18-based encoder with a U-Net decoder performed well in segmenting lung regions affected by various conditions, making it suitable for segmentation tasks.
- The customized CNN model achieved strong classification results, benefiting from various data augmentation techniques.

The VGG19-based model's superior performance in classifying COVID-19 can be attributed to its transfer learning strategy, while the ResNet18-U-Net architecture was optimized for biomedical image segmentation, excelling at feature extraction. The customized CNN model in Matlab offered flexibility in classification, achieving robust results across all metrics.

5. Discussion and Future Work

5.1. Discussion

This study explored the application of deep learning models, including VGG19, ResNet18-based encoder with a U-Net inspired decoder, and a Customized CNN model in the classification and segmentation of COVID-19 and related lung diseases using medical imaging. The performance of each model demonstrates the effectiveness of deep learning techniques in medical diagnostics, but also highlights certain challenges and opportunities for improvement. Below, we present a more detailed discussion of the models' limitations and the challenges faced during development.

VGG19 Model

The VGG19 model achieved high validity in classifying COVID-19 and other lung conditions, demonstrating the potential of transfer learning in medical imaging. Its success is largely due to the ability to leverage pre-trained layers to capture complex features. However, a significant limitation observed was the model’s tendency to overfit on the training data, which can be attributed to the restricted diversity of the dataset. This overfitting suggests that the model's generalization ability could be compromised when applied to unseen data. Additionally, VGG19’s computational complexity can hinder real-time or large-scale deployment in clinical settings. Increasing the dataset size, especially by including more diverse samples, and enhancing data augmentation techniques (e.g., brightness adjustment, rotation, scaling) would help mitigate this issue. Furthermore, regularization techniques like L2 regularization, dropout, and batch normalization will prevent overfitting and improve generalization.

ResNet18-Based Encoder with U-Net Inspired Decoder

The ResNet18-based encoder with U-Net inspired decoder was highly effective for biomedical image segmentation tasks, particularly when identifying lung regions affected by diseases like COVID-19 and viral pneumonia. The combination of ResNet18 for feature extraction and U-Net for segmentation enhanced the model’s ability to precisely capture both global and local structures within the images. However, a notable challenge was the model’s high computational cost, which makes it less suitable for scenarios with limited resources or where real-time predictions are needed. To improve its applicability, future work should focus on optimizing computational efficiency and validating the model on additional biomedical datasets to ensure robustness across a broader range of medical conditions. Testing it on more complex or unseen medical images would further enhance its real-world performance.

Customized CNN Model (Matlab)

The Customized CNN model developed in Matlab also delivered promising results, but its limitations were more evident when compared to the transfer learning-based architectures. While the customized CNN performed well in classification tasks, the absence of pre-trained layers, as used in VGG19 and ResNet18, limits its capacity to generalize across more complex datasets. Additionally, the analysis of the training-validation accuracy/loss graphs and confusion matrices revealed areas for further optimization, particularly in the management of class imbalances and feature extraction. Improving the model’s architecture and experimenting with deeper convolutional layers could lead to enhanced performance. Moreover, the model could benefit from more advanced data augmentation and regularization techniques to improve robustness across various datasets.

Overall, these deep learning models underscore the potential for AI-driven diagnostics in medical imaging. However, it is critical to address the challenges of generalization, overfitting, and computational cost to maximize the models' clinical applicability.

5.2. Challenges in Model Development

In the process of model development, several key challenges were identified:

1. **Dataset Limitations:** The relatively small and less diverse dataset posed challenges for generalization. Class imbalance was particularly noticeable, which could skew the model's performance toward overrepresented classes, such as healthy individuals or common lung diseases like pneumonia. More comprehensive datasets and synthetic data generation techniques would help improve performance on underrepresented classes like COVID-19.
2. **Overfitting Issues:** Overfitting, especially in the VGG19 model, was a significant concern. While data augmentation helped, more sophisticated techniques, such as adversarial training or contrastive learning, could offer more robust solutions. Additionally, cross-validation strategies were employed to better estimate the model's true performance, but further work is needed to explore how different regularization and early stopping mechanisms can be combined for optimal results.
3. **Computational Constraints:** The high computational cost of training deep learning models, especially for U-Net-inspired architectures, required the use of advanced hardware. Reducing the model's footprint via model pruning or quantization techniques could enhance its usability in real-time clinical environments.

5.3. Future Work

Based on the challenges and findings of this study, several directions for future research are proposed to enhance model performance and reliability:

VGG19 Model

Future research should prioritize expanding the dataset to improve the model’s generalizability. This could involve curating a larger set of X-ray images from a wider demographic and geographical spectrum to ensure that the model can handle more diverse cases. Additionally, exploring alternative deep learning architectures such as InceptionNet or EfficientNet could yield improvements in both accuracy and computational efficiency. Further development of data augmentation techniques that target underrepresented classes could also significantly enhance the model's performance in real-world clinical settings. Finally, additional regularization methods (e.g., label smoothing, dropout) and ensemble learning approaches should be tested to further prevent overfitting.

ResNet18-Based Encoder with U-Net Inspired Decoder

For the ResNet18-based U-Net model, future research should focus on optimizing the model's computational efficiency for real-time use. This could involve implementing model compression techniques like distillation or pruning to reduce resource consumption without sacrificing performance. Furthermore, applying the model to other types of medical imaging (e.g., CT scans, MRI) could validate its versatility across different domains of medical diagnostics. Collaborating with medical professionals to create user-friendly interfaces that allow for practical, real-time use of these models in clinical environments would be another beneficial step.

Customized CNN Model (Matlab)

Future work on the Customized CNN should focus on exploring more complex architectures, such as deeper networks or hybrid models that combine CNNs with attention mechanisms to improve feature detection and classification. Additionally, improving the model's generalization by testing on larger datasets and applying more aggressive regularization techniques (e.g., early stopping, weight decay) will help address overfitting. Lastly, comparing the model against newer architectures could offer insights into further performance improvements.

In conclusion, this research highlights the potential of deep learning models in the field of medical imaging, particularly for diagnosing COVID-19 and associated lung diseases. While promising, these models still face limitations, particularly in terms of generalization, computational efficiency, and handling diverse clinical data. Addressing these challenges through larger datasets, more robust architectures, and computational optimizations will be crucial in refining these models for real-world use.

6. Conclusion

This study explored the application of various deep learning models, including VGG19, a ResNet18-based encoder with a U-Net inspired decoder, and a Customized CNN model, for the classification and segmentation of COVID-19 and related lung diseases using chest X-rays and biomedical images. The results from all three models demonstrate the significant potential of deep learning in medical imaging, particularly for tasks that require rapid and accurate diagnoses, such as COVID-19 detection.

The VGG19 model achieved high accuracy, precision, specificity, and F1 score, proving that transfer learning and data augmentation techniques are effective for medical image classification. Its success underscores the potential of deep learning models to assist in clinical settings, particularly for urgent health conditions where timely diagnosis is critical. Despite these achievements, future work should focus on expanding datasets and exploring alternative architectures to further improve the model’s performance and generalization ability.

Similarly, the ResNet18-based encoder and U-Net inspired decoder demonstrated its effectiveness in biomedical image segmentation tasks, outperforming traditional methods. This model's success in segmenting lung diseases highlights its potential application in diverse medical imaging tasks. Future research should focus on testing the model on a broader range of biomedical datasets, optimizing its computational efficiency, and developing user-friendly interfaces for real-world clinical deployment.

The Customized CNN model, developed using Matlab, also delivered promising results in classifying lung diseases, achieving high accuracy and demonstrating its suitability for clinical applications. However, future efforts should involve testing the model with larger and more diverse datasets, and exploring advanced data augmentation and regularization techniques to further enhance its performance.

In conclusion, the findings of this study emphasize the practical benefits and potential of deep learning models in medical image analysis and diagnostics. With continued research focused on optimizing model architectures, expanding datasets, and improving generalization, these models can play a pivotal role in improving the quality of healthcare by enabling faster, more accurate, and reliable diagnoses in real-world clinical environments.

References

[1] Punn, N. S., & Agarwal, S. (2021). Automated diagnosis of COVID-19 with limited posteroanterior chest X-ray images using fine-tuned deep neural networks. Applied Intelligence, 51(5), 2689-2702.

[2] Srinivas, K., Gagana Sri, R., Pravallika, K., Nishitha, K., & Polamuri, S. R. (2024). COVID-19 prediction based on hybrid Inception V3 with VGG16 using chest X-ray images. Multimedia Tools and Applications, 83(12), 36665-36682.

[3] Govindarajan, S., & Swaminathan, R. (2021). Differentiation of COVID-19 conditions in planar chest radiographs using optimized convolutional neural networks. *Applied Intelligence*, *51*(5), 2764-2775.

[4] Wang, Y., Kang, H., Liu, X., & Tong, Z. (2020). Combination of RT-qPCR testing and clinical features for diagnosis of COVID-19 facilitates management of SARS-CoV-2 outbreak. *Journal of medical virology*, *92*(6), 538.

[5] Adhikari, S. P., Meng, S., Wu, Y. J., Mao, Y. P., Ye, R. X., Wang, Q. Z., ... & Zhou, H. (2020). Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review. *Infectious diseases of poverty*, *9*, 1-12.

[6] Pan, Y., Li, X., Yang, G., Fan, J., Tang, Y., Zhao, J., ... & Li, Y. (2020). Serological immunochromatographic approach in diagnosis with SARS-CoV-2 infected COVID-19 patients. *Journal of Infection*, *81*(1), e28-e32.

[7] Spicuzza, L., Montineri, A., Manuele, R., Crimi, C., Pistorio, M. P., Campisi, R., ... & Crimi, N. (2020). Reliability and usefulness of a rapid IgM-IgG antibody test for the diagnosis of SARS-CoV-2 infection: A preliminary report. *The Journal of infection*, *81*(2), e53.

[8] Porte, L., Legarraga, P., Vollrath, V., Aguilera, X., Munita, J. M., Araos, R., ... & Weitzel, T. (2020). Evaluation of a novel antigen-based rapid detection test for the diagnosis of SARS-CoV-2 in respiratory samples. *International Journal of Infectious Diseases*, *99*, 328-333.

[9] Nayak, S. R., Nayak, D. R., Sinha, U., Arora, V., & Pachori, R. B. (2022). An efficient deep learning method for detection of COVID-19 infection using chest X-ray images. *Diagnostics*, *13*(1), 131.

[10] Gupta, K., & Bajaj, V. (2023). Deep learning models-based CT-scan image classification for automated screening of COVID-19. Biomedical Signal Processing and Control, 80, 104268.

[11] Barshooi, A. H., & Amirkhani, A. (2022). A novel data augmentation based on Gabor filter and convolutional deep learning for improving the classification of COVID-19 chest X-Ray images. *Biomedical Signal Processing and Control*, *72*, 103326.

[12] Basu, S., Mitra, S., & Saha, N. (2020, December). Deep learning for screening covid-19 using chest x-ray images. In *2020 IEEE symposium series on computational intelligence (SSCI)* (pp. 2521-2527). IEEE.

[13] Che Azemin, M. Z., Hassan, R., Mohd Tamrin, M. I., & Md Ali, M. A. (2020). COVID-19 deep learning prediction model using publicly available radiologist-adjudicated chest x-ray images as training data: preliminary findings. *International Journal of Biomedical Imaging*, *2020*(1), 8828855.

[14] Aslan, M. (2022). Derin Öğrenme Tabanlı Otomatik Beyin Tümör Tespiti. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, *34*(1), 399-407.

[15] Erdoğan, A. M. N., Öztürk, T., & Talo, M. (2022). Yeni bir Evrişimsel Sinir Ağı Modeli Kullanarak Bilgisayarlı Tomografi Görüntülerinden Akciğer Kanseri Tespiti. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, *34*(2), 795-802.

[16] Khan, A. I., Shah, J. L., & Bhat, M. M. (2020). CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. *Computer methods and programs in biomedicine*, *196*, 105581.

[17] Narin, A., Kaya, C., & Pamuk, Z. (2021). Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *Pattern Analysis and Applications*, *24*, 1207-1220.

[18] Apostolopoulos, I. D., & Mpesiana, T. A. (2020). Covid-19: automatic detection from x-ray images

utilizing transfer learning with convolutional neural networks. *Physical and engineering sciences in medicine*, *43*, 635-640.

[19] Wang, L., Lin, Z. Q., & Wong, A. (2020). Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific reports*, *10*(1), 19549.

[20] Kroft, L. J., van der Velden, L., Girón, I. H., Roelofs, J. J., de Roos, A., & Geleijns, J. (2019). Added value of ultra–low-dose computed tomography, dose equivalent to chest x-ray radiography, for diagnosing chest pathology. *Journal of thoracic imaging*, *34*(3), 179-186.

[21] Damilakis, J., Adams, J. E., Guglielmi, G., & Link, T. M. (2010). Radiation exposure in X-ray-based imaging techniques used in osteoporosis. *European radiology*, *20*, 2707-2714.

[22] Su, Q., Kloukinas, C., & Garcez, A. D. A. (2024, June). FocusLearn: Fully-Interpretable, High-Performance Modular Neural Networks for Time Series. In *2024 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

[23] Fayemiwo, M. A., Olowookere, T. A., Arekete, S. A., Ogunde, A. O., Odim, M. O., Oguntunde, B. O., ... & Kayode, A. A. (2021). Modeling a deep transfer learning framework for the classification of COVID-19 radiology dataset. *PeerJ Computer Science*, *7*, e614.

[24] Cao, Z., Huang, J., He, X., & Zong, Z. (2022). BND-VGG-19: A deep learning algorithm for COVID-19 identification utilizing X-ray images. *Knowledge-Based Systems*, *258*, 110040.

[25] Xu, X., Jiang, X., Ma, C., Du, P., Li, X., Lv, S., ... & Li, L. (2020). A deep learning system to screen novel coronavirus disease 2019 pneumonia. *Engineering*, *6*(10), 1122-1129.

[26] El Asnaoui, K., & Chawki, Y. (2021). Using X-ray images and deep learning for automated detection of coronavirus disease. *Journal of Biomolecular Structure and Dynamics*, *39*(10), 3615-3626.

[27] Elzeki, O. M., Shams, M., Sarhan, S., Abd Elfattah, M., & Hassanien, A. E. (2021). COVID-19: a new deep learning computer-aided model for classification. *PeerJ Computer Science*, *7*, e358.

[28] Mahmud, T., Rahman, M. A., & Fattah, S. A. (2020). CovXNet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization. *Computers in biology and medicine*, *122*, 103869.

[29] Song, Y., Zheng, S., Li, L., Zhang, X., Zhang, X., Huang, Z., ... & Yang, Y. (2021). Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. *IEEE/ACM transactions on computational biology and bioinformatics*, *18*(6), 2775-2780..

[30] Rahaman, M. M., Li, C., Yao, Y., Kulwa, F., Rahman, M. A., Wang, Q., ... & Zhao, X. (2020). Identification of COVID-19 samples from chest X-Ray images using deep learning: A comparison of transfer learning approaches. *Journal of X-ray Science and Technology*, *28*(5), 821-839.

AITA
ARTIFICIAL INTELLIGENCE THEORY AND APPLICATIONS

■ RESEARCH ARTICLE

Intrusion Detection on CSE-CIC-IDS2018 Dataset Using Machine Learning Methods

Halil İbrahim Coşar[a†], **Çağrı Arısoy**[b], **Hasan Ulutaş**[b]

[a] Department of Electric-Electronic Engineering, Yozgat Bozok University, Yozgat, Türkiye
[b] Department of Computer Engineering, Yozgat Bozok University, Yozgat, Türkiye
† halil.cosar@bozok.edu.tr, corresponding author

RECEIVED SEPTEMBER 20, 2024
ACCEPTED SEPTEMBER 30, 2024

CITATION Coşar, H. İ., Arısoy, Ç., & Ulutaş. H. (2024). Intrusion detection on CSE-CIC-IDS2018 dataset using machine learning methods. *Artificial Intelligence Theory and Applications, 4(2*), 143-154.

Abstract

Over the past few decades, the significance of computer and information security has grown exponentially, driven by the escalating frequency and sophistication of cyber threats. Despite the rapid advancements in both intrusion techniques and security technologies, many organizations continue to rely on outdated cybersecurity strategies, leaving them vulnerable to increasingly complex cyberattacks. Conventional defenses, such as basic firewalls and signature-based detection systems, are often insufficient against modern attackers who use advanced methods, including zero-day exploits and polymorphic malware, to evade detection. Government web servers, which house vast amounts of sensitive citizen data, are especially attractive targets for malicious actors. In response to these evolving threats, the deployment of an Intrusion Detection System (IDS) has become a critical component in securing network infrastructures, providing an essential layer of defense against unauthorized access and data breaches. This study explores the efficacy of six distinct machine learning-based classification methods; Random Forest, Gradient Boosting, XGBoost, CatBoost, Logistic Regression, and LightGBM each selected for its particular strengths in handling complex, high-dimensional data. These algorithms were applied to a comprehensive dataset to detect malicious activities, with a focus on achieving high accuracy and robustness in classification performance. Remarkably, all six models demonstrated substantial effectiveness, achieving accuracy rates as high as 0.98 and AUC values reaching 1.00, underscoring their potential in enhancing IDS capabilities. The results highlight the importance of leveraging advanced machine learning techniques in bolstering cybersecurity defenses, particularly in critical domains like government data protection, where precision and reliability are paramount.

Keywords: network intrusion, classification, cyber security, machine learning

1. Introduction

Computer and information security has become an increasingly significant issue over the past decades. While intrusion techniques and security protections have advanced rapidly, many organizations continue to rely on outdated cybersecurity measures. These traditional defences are often inadequate against modern cyberattacks, which use sophisticated methods to bypass them. Government web servers, which store sensitive

Artificial Intelligence Theory and Applications, ISSN: 2757-9778. ISBN: 978-605-69730-2-4 © 2024 İzmir Bakırçay University

information about citizens, are particularly attractive targets for hackers [1]. Today, an Intrusion Detection System (IDS) is an essential defence mechanism critical for safeguarding important networks against intrusions [2]. IDSs can be categorized into two types: anomaly-based and signature-based. Anomaly-based IDSs operate by creating a model of normal system behaviour and identifying any deviations from this baseline. In contrast, signature-based IDSs rely on a database of known attack signatures to recognize malicious activities [3]. In the commercial sector, signature-based IDSs are commonly employed. However, anomaly-based IDSs have the advantage of being able to detect previously unknown attacks. Despite this, anomaly-based IDSs typically suffer from low detection rates and high false positive rates. To improve the detection of new attacks, adaptive and efficient Machine Learning (ML) and Deep Learning (DL) algorithms are frequently utilized [4].

2. Related Work

Two recent public datasets, CICIDS2017 [5] and CSE-CIC-IDS2018 [6], are now available and include normal traffic as well as contemporary attack scenarios such as Heartbleed, Brute-force, Botnet, and Denial of Service (DoS). Although these datasets are accessible to the public, there has been limited use of them for evaluating, testing, and fine-tuning real-time IDS deployments.

Atefinia and Ahmadi [1] propose a multi-architectural modular deep neural network model aimed at enhancing anomaly-based intrusion detection systems by reducing the false-positive rate. This model includes a feed-forward module, a stack of restricted Boltzmann machine modules, and two recurrent modules, with their output weights combined in an aggregator module to make the final decision. Experiments using the CSE-CIC-IDS2018 dataset show significant improvements in detecting specific network attacks, achieving up to 100% accuracy for certain network-level attacks compared to existing methods. The models developed in this study can be effectively used in IDS to generate alerts or prevent new attacks. This deep neural network model offers a promising solution to the limitations of traditional signature-based intrusion detection systems by utilizing machine learning techniques to detect network attacks without relying solely on predefined signatures. In Basnet et al. [7] deep learning algorithms have demonstrated significant potential in network intrusion detection, as evidenced. Researchers assessed the effectiveness of several state-of-the-art deep learning frameworks, including Keras, TensorFlow, Theano, fast.ai, and PyTorch, in identifying and classifying network intrusion traffic. Using the CSE-CIC-IDS2018 dataset to evaluate these frameworks, fast.ai, a PyTorch wrapper, achieved the highest accuracy, approximately 99%, with low false positive and false negative rates in detecting and classifying various types of network intrusions. This high level of accuracy underscores the potential of deep learning frameworks in effectively identifying and categorizing network attacks. The results strongly support the effectiveness and utility of deep learning frameworks in network intrusion detection, emphasizing the importance of leveraging these techniques to enhance cybersecurity measures and effectively combat evolving cyber threats. Another paper evaluated two traditional training algorithms for Hidden Markov Models (HMM), Baum Welch (BW) and Viterbi Training (VT), using three standard initialization techniques: uniform, random, and count-based. The performance of the HMM was analysed based on detecting all states (AS), the current state (CS), and the next state (NS) given an observation sequence. The count-based initialization technique outperformed the uniform and random techniques in detecting AS and CS, achieving about 97.5% and 97.0% accuracy for AS prediction using BW and VT, respectively. For CS detection, the performance was similar to AS detection, with a slight decrease of about 0.2%. Predicting NS had an accuracy of around 65% for both uniform and random initialization techniques with BW and VT. The study found no significant improvement with increasing the window sample size, and the training techniques can be practically implemented by connecting the output of an IDS or a database storing alerts to an HMM [4]. In the other study explored the inter-dataset generalization of supervised machine learning methods for intrusion detection, aiming to differentiate between benign and various types of malicious network traffic. Classification benchmarks were established using two labelled datasets, CIC-IDS2017 and CSE-CIC-IDS2018, which include attack classes such as DoS, DDoS, infiltration, and botnet. Twelve supervised learning algorithms from different families were compared. The research revealed that high generalization within a dataset does not necessarily translate to high generalization across different datasets, especially for attack types like DoS/SSL and botnet. The trained models failed to maintain high classification performance when tested on new but related samples without additional training. These findings challenge the assumption that strong intra-dataset performance guarantees strong inter-dataset performance. Further investigation is needed to understand the limitations and develop solutions to enhance inter-dataset generalization in supervised ML-based intrusion detection systems [8]. Another paper presented a comparative analysis of deep learning methods for intrusion detection, specifically examining deep discriminative models and generative unsupervised models. Seven different deep learning techniques were evaluated: recurrent neural networks (RNNs), deep neural networks (DNNs), restricted Boltzmann machines (RBMs), deep belief networks (DBNs), convolutional neural networks (CNNs), deep Boltzmann machines (DBMs), and deep autoencoders. The evaluation was conducted using two novel datasets, CSE-CIC-IDS2018 and Bot-IoT, and was based on three primary performance metrics: false alarm rate, accuracy, and detection rate. The goal of the study was to assess the effectiveness of these deep learning models in various intrusion detection scenarios, offering insights into their performance for both binary and multiclass classification tasks. The findings are crucial for advancing cybersecurity measures by employing sophisticated deep learning techniques, thereby enhancing the accuracy and efficacy of intrusion detection systems in identifying cyber threats. Deep autoencoders exhibited the highest accuracy on both the CSE-CIC-IDS2018 and Bot-IoT datasets, with accuracy rates of 97.372 and 98.394, respectively. These results were achieved using a configuration of 100 hidden nodes and a learning rate of 0.5 [9]. Fitni and Ramli employed ensemble learning, which combined logistic regression, decision trees, and gradient boosting, to increase the performance of intrusion detection systems This method harnessed the strengths of each classifier to enhance detection accuracy, minimize false alarms, and improve the identification of unknown attacks. Feature selection techniques were used to pinpoint the most critical data features for intrusion detection. Using Spearman’s rank correlation coefficient, 23 out of 80 features were selected, enhancing the model's efficiency by concentrating on the most informative features. The proposed model achieved high performance on the CSE-CIC-IDS2018 dataset, attaining an accuracy of 98.8%, precision of 98.8%, recall of 97.1%, and an F1 score of 97.9%. These results underscore the effectiveness of ensemble learning and feature selection in improving anomaly-based intrusion detection systems, significantly enhancing detection capabilities, reducing false alarms, and bolstering overall network security within organizational information systems [10]. Kanimozhi and Jacob proposed a system which applies AI to the CSE-CIC-IDS2018 dataset and achieves outstanding performance metrics: 99.97% accuracy, an average area under the ROC curve of 0.999, and a low false positive rate of 0.001. These results highlight the system's high accuracy and precision in detecting botnet attacks. Its effectiveness in identifying botnet attacks underscores its potential to improve security in financial sectors and banking services, where such threats are particularly serious. This demonstrates the practical importance and applicability of AI-based intrusion detection systems in protecting critical systems and data. Additionally, the system's scalability allows for deployment across multiple machines, making it suitable for various applications such as network traffic analysis, cyber-physical system traffic data analysis, and real-time network traffic monitoring. This versatility enhances its relevance and utility in diverse cybersecurity contexts [11]. In another study, six machine learning models were implemented using the CSE-CIC-IDS2018 dataset. Data sampling techniques, such as the Synthetic Minority Oversampling Technique (SMOTE), were applied to increase the representation of minority classes and enhance detection rates for less common intrusions. The experimental results indicated that the implemented models achieved a high level of accuracy compared to recent studies. Using a sampled dataset led to an increase in the average accuracy of the models by between 4.01% and 30.59% [12].

3. Materials and Methods

3.1. Dataset

The CSE-CIC-IDS2018 dataset contains network traffic data from various services and protocols, predominantly HTTPS and HTTP, along with others like SMTP, POP3, IMAP, SSH, and FTP. It includes numerous attack scenarios. The final dataset encompasses seven distinct attack scenarios: brute-force attacks, Heartbleed exploitation, botnet activity, DoS (Denial of Service), DDoS (Distributed Denial of Service), web-based attacks, and internal network infiltration. The attacking infrastructure is composed of 50 machines, while the targeted organization includes five departments, comprising 420 computers and 30 servers. The network traffic from this dataset was processed using the CICFlowMeter-V3 tool, extracting 80 features for training, such as the number of packets per second, specific TCP flag packet counts, and the standard deviation of packet sizes in a session [6].

3.2. Preprocessing

In the data set one file includes 84 features and this file was not processed because files with an equal number of features were processed in this study. Then, the intrusions within the CIC-IDS2018 training dataset were categorized into two traffic types: benign and attack. To streamline the experiments and ensure clarity, any data points containing Infinity or NaN values were excluded from the dataset, which also helped improve the quality of the input data for the models. In cases where text data was present, it was converted to float to ensure uniformity in the dataset and to facilitate the mathematical operations needed for machine learning models. Timestamps, which did not contribute significantly to the feature space, were removed from the dataset to avoid any potential bias in time-based patterns. Following this, the dataset underwent a normalization process using the StandardScaler technique. This approach scales the data such that it has a mean of 0 and a standard deviation of 1, which is often critical for models that are sensitive to the scale of features. Normalization helps ensure that features with varying ranges do not disproportionately influence the model's learning process, resulting in a more balanced and accurate performance. The pre-processed dataset was then split into training and validation sets in an 80-20 ratio, with 80% of the data allocated for training and 20% reserved for validation. The training set was employed to fit the machine learning models, while the validation set was used to evaluate the final model performance, ensuring that the models could generalize well to unseen data. To address the issue of class imbalance, an under-sampling technique was applied to the training set. This process involved reducing the number of samples in the majority class, which in this case was the benign traffic data, to match or closely match the minority class, representing the attack traffic. By randomly removing excess samples from the majority class, a more balanced dataset was created, which helped the models avoid overfitting to the dominant class and improved their ability to detect intrusions in the minority class. This step was crucial for enhancing model accuracy, particularly in imbalanced data scenarios where the majority class can overwhelm the learning algorithm.

3.3. Evaluation metrics

Various metrics are commonly used to assess and compare the performance of machine learning classifiers. The proposed model was evaluated using the following performance metrics.
Accuracy: Measures the percentage of correctly classified samples out of the total number of samples. The formula is:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Recall (Sensitivity): The ratio of correctly classified samples of a specific category (X) to the total samples of that category, indicating the system’s effectiveness in detecting anomalies.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

Precision: Represents the ratio of correctly predicted positive observations to the total predicted positives.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

F1 Score: The harmonic mean of precision and recall, accounting for both false positives and false negatives.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

where TP, FP, TN, and FN represent true positives, false positives, true negatives, and false negatives, respectively.

3.4. Classification

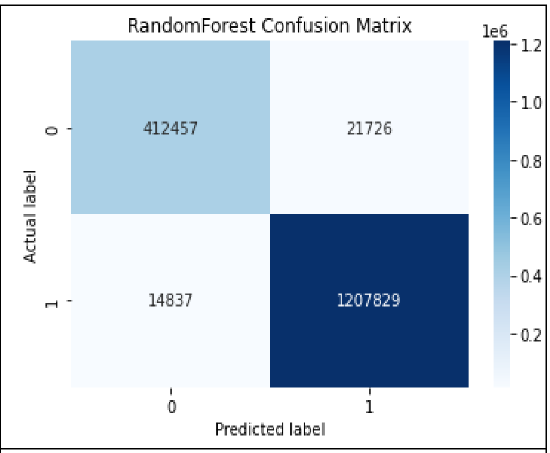
In this study, six different classification methods: random forest, gradient boosting, XGBoost, CatBoost, logistic regression, and LightGBM were implemented. Each of these methods was chosen for its unique strengths. Random forest reduces overfitting and handles noisy data well by constructing multiple decision trees. Gradient boosting incrementally improves performance by correcting errors from weak learners. XGBoost, an optimized version of GBM, offers faster performance and handles large datasets effectively. CatBoost excels with categorical data and requires less preprocessing. Logistic regression provides a simple yet powerful approach for linear relationships and is easily interpretable. LightGBM is optimized for large datasets and delivers high-speed performance with low memory usage. By using these diverse methods, it is aimed to explore various model structures and approaches to achieve optimal classification performance based on the dataset's characteristics. In the classification, the system being used is equipped with 64 GB of memory and is powered by two Intel(R) Xeon(R)

Silver 4114 CPUs, each running at 2.20 GHz. The server model is an HP Z6 G4, and it features an NVIDIA GeForce RTX 3090 Ti graphics card. The operating system is Windows 10 Pro for Workstations, and Python 3 is the language being used within the Jupyter Notebook framework.

4. Results and Discussion

While the results of used classification algorithm are analysed, three important visuals are used which are confusion matrix, ROC (Receiver Operating Characteristic) curve and learning curve. The ROC curve, learning curve, and confusion matrix are essential tools for evaluating classification models. The ROC curve plots the true positive rate (sensitivity) against the false positive rate, helping to assess a model's performance across different thresholds and its ability to distinguish between classes. The area under the ROC curve (AUC) is a key metric, where a higher value indicates better performance. The learning curve shows how a model's accuracy or error rate changes with varying amounts of training data, offering insights into whether the model is underfitting or overfitting and how it improves as it learns from more data. Finally, the confusion matrix provides a detailed breakdown of the model's predictions, showing true positives, true negatives, false positives, and false negatives, enabling a more granular understanding of classification accuracy and potential misclassifications. Together, these tools give a comprehensive view of a model's effectiveness, training behaviour, and areas for improvement.

RF	*precision*	*recall*	*f1-score*	*Support*
0	0.97	0.95	0.96	434183
1	0.98	0.99	0.99	1222666
accuracy			0.98	1656849
macro avg	0.97	0.97	0.97	1656849
weighted avg	0.98	0.98	0.98	1656849



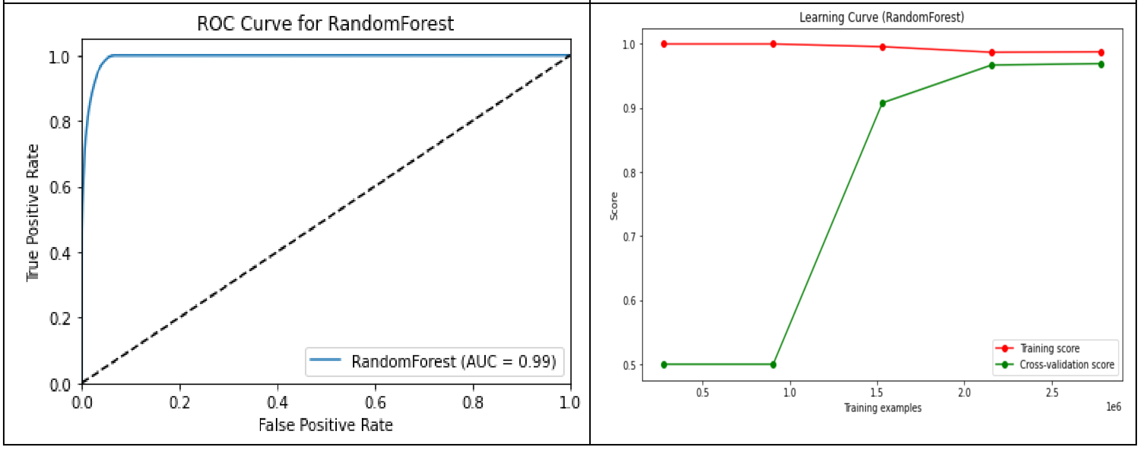
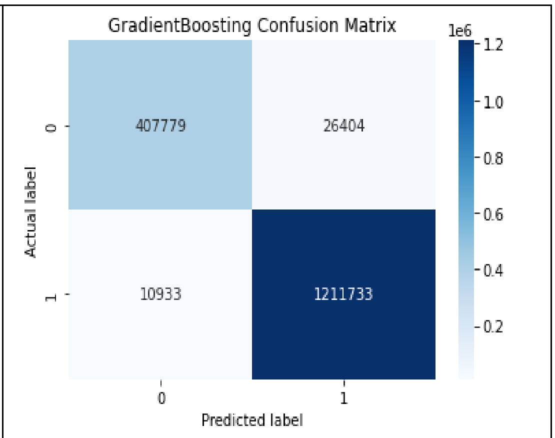


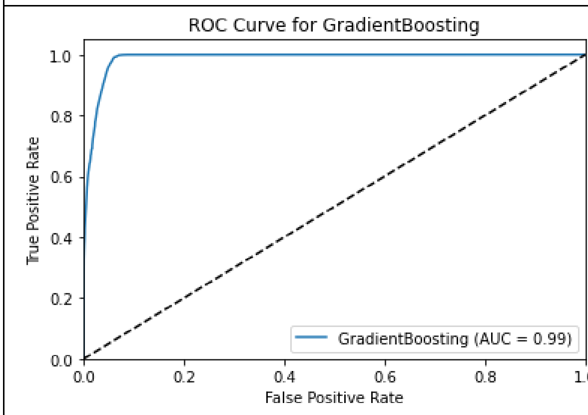
Figure 1. Random Forest classification report

The performance of the machine learning algorithms used in this study is illustrated in Figures 1-6. Based on these results, it is evident that all six techniques demonstrate exceptional performance on the given dataset, highlighting their suitability for network intrusion detection tasks. XGBoost, LightGBM, and CatBoost emerge as the top-performing models, achieving an impressive accuracy rate of 0.98 and an AUC score of 1.00, signifying near-perfect classification capabilities. These results suggest that these gradient-boosting-based methods are highly effective at distinguishing between normal and malicious network traffic, likely due to their advanced handling of complex interactions and non-linear relationships within the data.

Similarly, the Gradient Boosting and Random Forest algorithms also achieve strong performance, reaching an accuracy rate of 0.98 and an AUC value of 0.99. While slightly below the top-performing models, these results still demonstrate robust classification abilities, confirming their reliability in identifying potential intrusions. The success of these ensemble methods may be attributed to their ability to reduce overfitting and enhance model generalization by combining the predictions of multiple trees.

GB	*precision*	*recall*	*f1-score*	*support*
0	0.97	0.94	0.96	434183
1	0.98	0.99	0.98	1222666
accuracy			0.98	1656849
macro avg	0.98	0.97	0.97	1656849
weighted avg	0.98	0.98	0.98	1656849





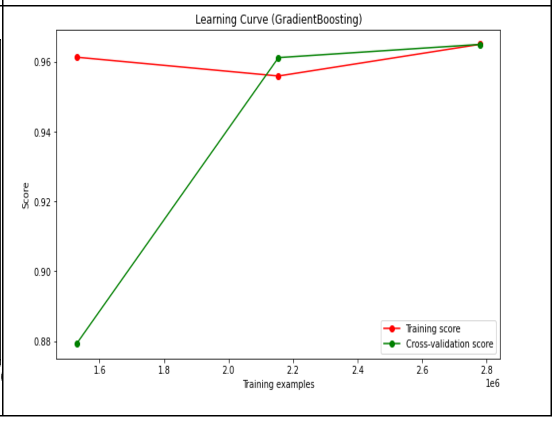
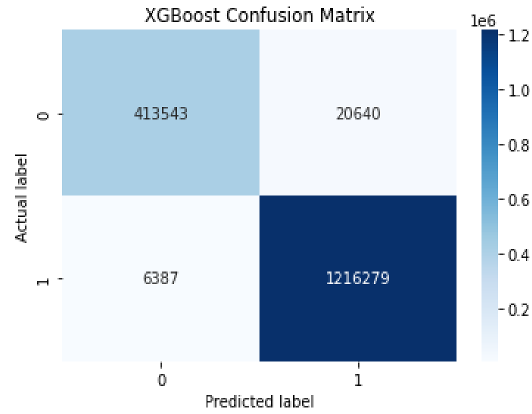
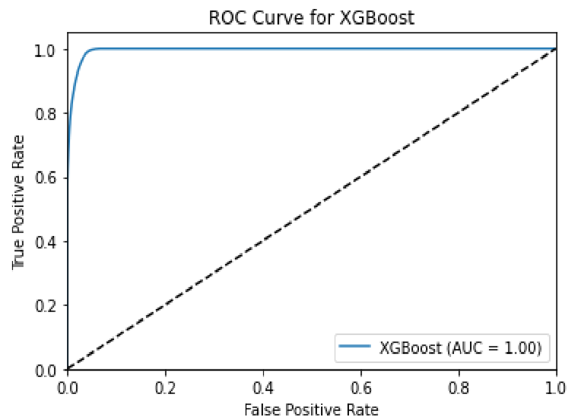


Figure 2. Gradient Boosting classification report

XGBoost	*precision*	*recall*	*f1-score*	*support*
0	0.98	0.95	0.97	434183
1	0.98	0.99	0.99	1222666
accuracy			0.98	1656849
macro avg	0.98	0.97	0.98	1656849
weighted avg	0.98	0.98	0.98	1656849





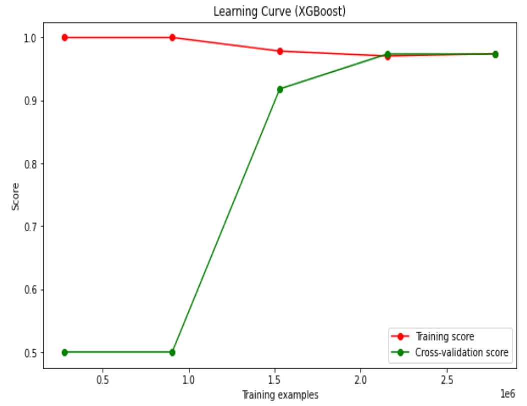
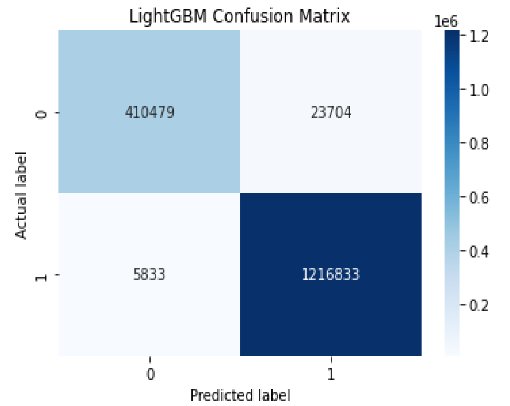
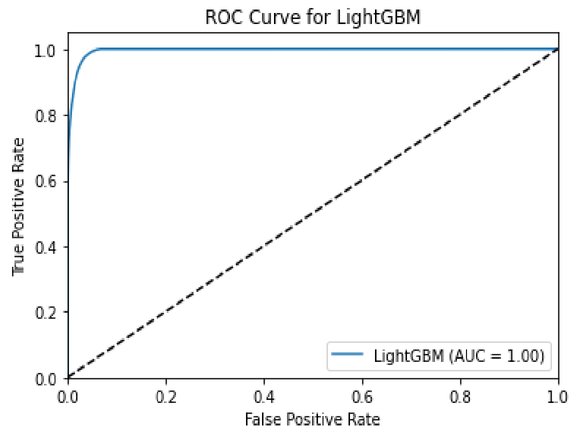


Figure 3. XGBoost classification report

LightGBM	*precision*	*recall*	*f1-score*	*support*
0	0.99	0.95	0.97	434183
1	0.98	1.00	0.99	1222666
accuracy			0.98	1656849
macro avg	0.98	0.97	0.98	1656849
weighted avg	0.98	0.98	0.98	1656849





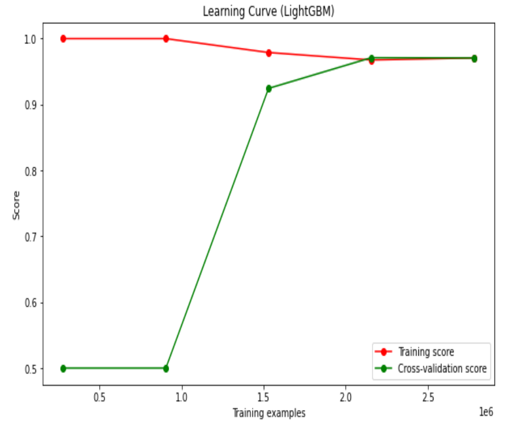
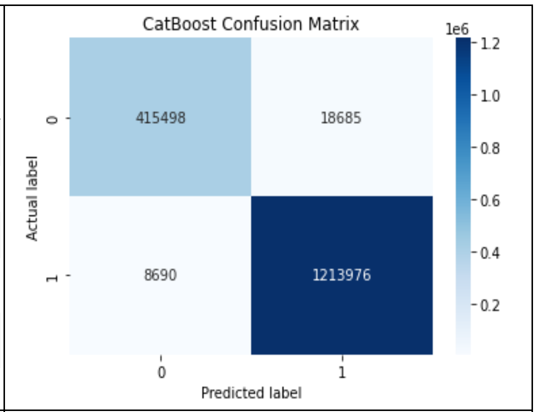


Figure 4. LightGBM classification report

CatBoost	*precision*	*recall*	*f1-score*	*support*
0	0.98	0.96	0.97	434183
1	0.98	0.99	0.99	1222666
accuracy			0.98	1656849
macro avg	0.98	0.97	0.98	1656849
weighted avg	0.98	0.98	0.98	1656849



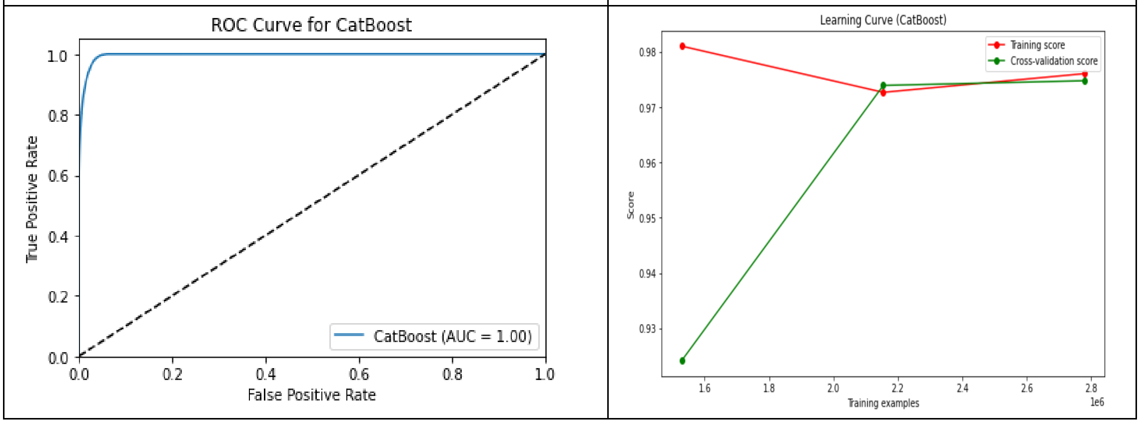
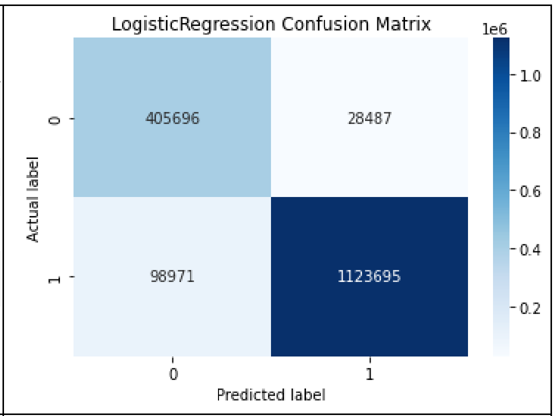


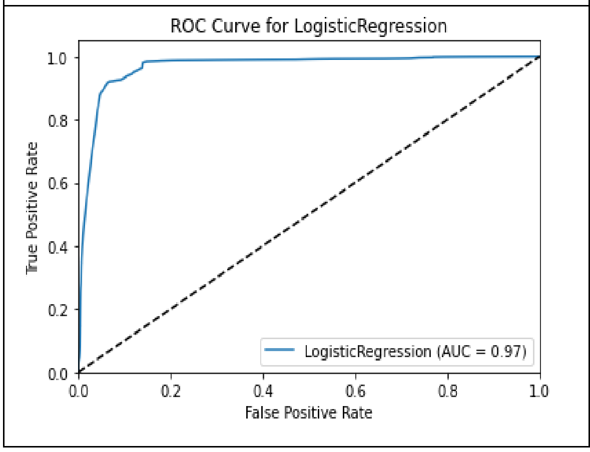
Figure 5. CatBoost classification report

In contrast, Logistic Regression, while comparatively less successful than the ensemble-based techniques, still delivers commendable results, with an accuracy of 0.92 and an AUC score of 0.97. Although it does not match the performance of the tree-based models, these results indicate that Logistic Regression remains a viable option for intrusion detection, particularly in scenarios where interpretability and simplicity are prioritized. Its lower performance could be due to its linear nature, which may limit its ability to capture more complex relationships in the data compared to non-linear models like gradient boosting or random forests.

In summary, while all the algorithms show strong performance, the results suggest that gradient-boosting-based methods, particularly XGBoost, LightGBM, and CatBoost, offer superior accuracy and AUC values, making them ideal for network intrusion detection. The relatively lower performance of Logistic Regression, although still effective, highlights the importance of algorithm selection based on the complexity and nature of the dataset. In conclusion, while all the algorithms demonstrate solid performance, gradient-boosting-based methods, specifically XGBoost, LightGBM, and CatBoost, stand out by providing the highest accuracy, making them particularly well-suited for network intrusion detection tasks. Although Logistic Regression performs adequately, its comparatively lower results emphasize the significance of choosing the right algorithm based on the dataset's complexity and characteristics.

LR	precision	recall	f1-score	support
0	0.80	0.93	0.86	434183
1	0.98	0.92	0.95	1222666
accuracy			0.92	1656849
macro avg	0.89	0.93	0.91	1656849
weighted avg	0.93	0.92	0.92	1656849





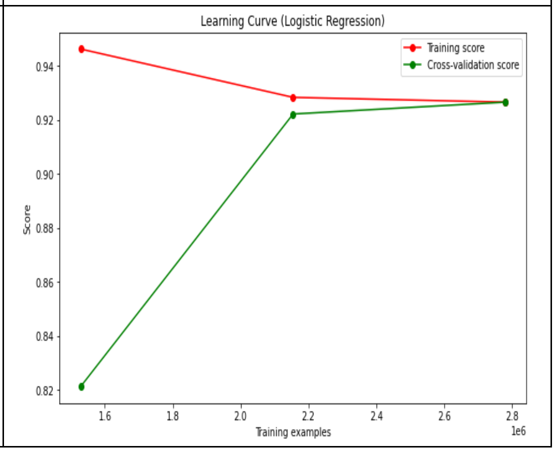


Figure 6. Logistic Regression classification report

During the classification process 5-fold cross validation was applied in order to evaluate the performance and generalization ability of machine learning models. The results are shown in figure 7. K-fold cross-validation bar chart provides a comparative visual representation of how well different classification algorithms performed on the dataset. The model with the longest bar was the most successful in classifying data consistently across all folds, while the algorithms with shorter bars were less accurate or consistent. This visual helps identify the strongest classification model, with attention to the differences in performance.

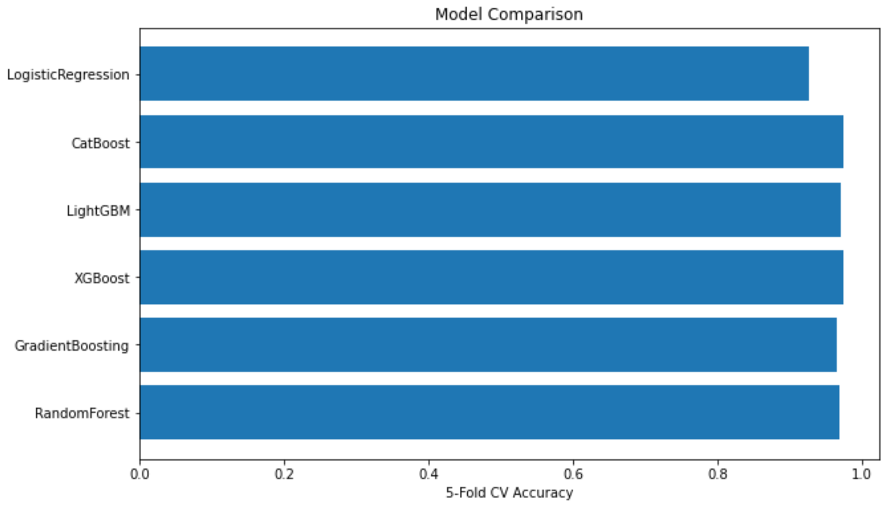


Figure 7. 5-fold cross validation result

5. Conclusion and Future Work

In this study, six different machine learning methods which are Random Forest, Gradient Boosting, XGBoost, CatBoost, Logistic Regression, and LightGBM are applied for detecting intrusions in network traffic, each demonstrating considerable potential in enhancing IDS. Our experimental results underscore the effectiveness of these algorithms, particularly when paired with appropriate preprocessing techniques. By reducing false positives for certain types of intrusions and achieving an accuracy rate of up to 98%, these methods offer promising alternatives to conventional detection systems. The performance we observed is not only competitive but also exceeds the benchmarks reported in much of the existing literature, highlighting the significance of integrating machine learning approaches for network security.

Despite the success of these models, there remain numerous opportunities for future research. One key direction would be to further refine feature extraction techniques to more accurately capture the characteristics of network traffic, particularly for anomaly-based intrusion detection systems. The integration of advanced feature engineering, or the use of deep learning-based automatic feature extraction, could potentially uncover hidden patterns in network data, further improving detection accuracy and reducing false alarms. Moreover, different types of datasets, including real-world network traffic from varied domains, could be explored using the methodology outlined in this research. This would provide a broader understanding of how these algorithms generalize across diverse environments and attack scenarios.

Another promising area for future work is the exploration of hybrid models that combine the strengths of multiple machine learning techniques, or the development of ensemble methods tailored specifically to network intrusion detection. Additionally, the impact of real-time data processing and online learning could be investigated to assess how well these models perform in dynamic environments where network conditions change frequently. Finally, further investigation into model interpretability and the ability to explain detection decisions will be crucial for fostering trust in machine learning-driven IDS systems, especially in high-stakes domains like government, healthcare, and financial networks.

By continuing to build upon the findings of this study, future research has the potential to significantly advance the capabilities of IDS systems, leading to more robust and adaptive network security solutions capable of defending against increasingly sophisticated cyber threats.

References

[1] Atefinia, R., & Ahmadi, M. (2021). Network intrusion detection using multi-architectural modular deep neural network. Journal of Supercomputing, 77(4), 3571–3593. https://doi.org/10.1007/S11227-020-03410-Y/FIGURES/14

[2] Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. ICISSP 2018 - Proceedings of the 4th International Conference on Information Systems Security and Privacy, 2018-January, 108–116. https://doi.org/10.5220/0006639801080116

[3] Hussein, S. M. (2016). Performance Evaluation of Intrusion Detection System Using Anomaly and Signature Based Algorithms to Reduction False Alarm Rate and Detect Unknown Attacks. 2016 International Conference on Computational Science and Computational Intelligence (CSCI), 1064–1069. https://doi.org/10.1109/CSCI.2016.0203

[4] Chadza, T., Kyriakopoulos, K. G., & Lambotharan, S. (2019). Contemporary Sequential Network Attacks Prediction using Hidden Markov Model. 2019 17th International Conference on Privacy, Security and Trust, PST 2019 - Proceedings. https://doi.org/10.1109/PST47121.2019.8949035

[5] IDS 2017 | Datasets | Research | Canadian Institute for Cybersecurity | UNB. (n.d.). Retrieved May 29, 2024, from https://www.unb.ca/cic/datasets/ids-2017.html

[6] IDS 2018 | Datasets | Research | Canadian Institute for Cybersecurity | UNB. (n.d.). Retrieved May 29, 2024, from https://www.unb.ca/cic/datasets/ids-2018.html

[7] Basnet, R., Johnson, C., Basnet, R. B., Shash, R., Walgren, L., & Doleck, T. (n.d.). Towards Detecting and Classifying Network Intrusion Traffic Using Deep Learning Frameworks. Researchgate.NetRB Basnet, R Shash, C Johnson, L Walgren, T DoleckJ. Internet Serv. Inf. Secur., 2019•researchgate.Net. https://doi.org/10.22667/JISIS.2019.11.30.001

[8] D'hooge, L., Wauters, T., Volckaert, B., & De Turck, F. (2020). Inter-dataset generalization strength of supervised machine learning methods for intrusion detection. Journal of Information Security and Applications, 54, 102564. https://doi.org/10.1016/J.JISA.2020.102564

[9] Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. Journal of Information Security and Applications, 50, 102419. https://doi.org/10.1016/J.JISA.2019.102419

[10] Fitni, Q. R. S., & Ramli, K. (2020). Implementation of ensemble learning and feature selection for performance improvements in anomaly-based intrusion detection systems. Proceedings - 2020 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology, IAICT 2020, 118–124. https://doi.org/10.1109/IAICT50021.2020.9172014

[11] Kanimozhi, V., & Prem Jacob, T. (2019). Artificial intelligence based network intrusion detection with hyper-parameter optimization tuning on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing. Proceedings of the 2019 IEEE International Conference on Communication and Signal Processing, ICCSP 2019, 33–36. https://doi.org/10.1109/ICCSP.2019.8698029

[12] Karatas, G., Demir, O., & Sahingoz, O. K. (2020). Increasing the Performance of Machine Learning-Based IDSs on an Imbalanced and Up-to-Date Dataset. IEEE Access, 8, 32150–32162. https://doi.org/10.1109/ACCESS.2020.2973219

AITA
ARTIFICIAL INTELLIGENCE THEORY AND APPLICATIONS

■ REVIEW ARTICLE

Prediction of Lung Cancer with Fuzzy Logic Methods: A Systematic Review

Beyza Aslan [a†], **Ouranıa Areta Hızıroğlu** [a]

[a] Department of Management Information Systems, İzmir Bakırçay University, İzmir, Türkiye
† beyza.aslan@bakircay.edu.tr, corresponding author

RECEIVED SEPTEMBER 13, 2024
ACCEPTED SEPTEMBER 29, 2024

CITATION Aslan, B. & Areta Hızıroğlu, O. (2024). Prediction of lung cancer with fuzzy logic methods: A systematic review. *Artificial Intelligence Theory and Applications, 4*(2), 153-192.

Abstract

According to the World Health Organization (WHO), lung cancer is the primary cause of cancer-related deaths worldwide and is known to have the highest mortality rate among both men and women. Early and accurate detection of lung cancer can lead to better treatments and outcomes. Different methods can be used to diagnose a complex and uncertain disease, such as lung cancer, and fuzzy logic is one of these methods. The challenge of diagnosing lung cancer nodules, coupled with the high mortality rate of lung cancer, underscores the significance of using fuzzy logic. Fuzzy logic offers a problem-solving approach that relies on logical rules and if-then statements, incorporating human experience. There are many studies in the literature on the diagnosis of lung cancer with fuzzy logic approaches, and it is important to examine these studies to provide a general framework on this subject. Therefore, this systematic review aims to synthesize and evaluate the current evidence on the application of fuzzy logic methods in lung cancer prediction and diagnosis, and thus can provide a guide to researchers and decision makers who want to work in this field. The study followed the PRISMA guidelines for systematic reviews, ensuring a structured and transparent approach to the research process. Scopus, Web of Science (WoS), PubMed, and IEEE Explore databases were searched to find relevant studies, and appropriate studies were carefully reviewed. The inclusion and exclusion criteria were clearly defined, and the analysis process was performed independently. Out of 222 initially identified studies, 51 met the inclusion criteria and were analyzed in depth. The most commonly used fuzzy logic methods were Fuzzy Rule-Based Systems, Fuzzy C-Means Clustering, and Fuzzy Inference Systems. Studies reported accuracy rates ranging from 85% to 98% in lung cancer prediction and diagnosis. Hybrid models combining fuzzy logic with other machine learning techniques showed particularly promising results. Fuzzy logic methods demonstrate significant potential in improving the accuracy of lung cancer prediction and diagnosis. However, further research is needed to standardize approaches and validate these methods in large-scale clinical settings. The integration of fuzzy logic with other artificial intelligence techniques presents a promising direction for future developments in lung cancer diagnostics.

Keywords: fuzzy logic, lung cancer, prediction, diagnosis, systematic review

---

1. Introduction

Cancer is on the rise worldwide due to environmental factors, nutritional conditions, and genetic factors and has become the leading cause of death due to inadequate treatment

Artificial Intelligence Theory and Applications, ISSN: 2757-9778. ISBN: 978-605-69730-2-4 © 2024 İzmir Bakırçay University

and diagnosis. The WHO states that cancer is one of the leading causes of death worldwide, and recent data on cancer mortality rates show that lung cancer is the most common cause of death among cancer types with 2.48 million cases worldwide (Figure 1) [1]. Lung cancer remains a major global health problem, causing morbidity and mortality worldwide. Early and accurate detection is crucial for improving patient outcomes and survival rates [2]. This pressing need has driven researchers and healthcare professionals to explore innovative approaches in diagnostic technologies [3].

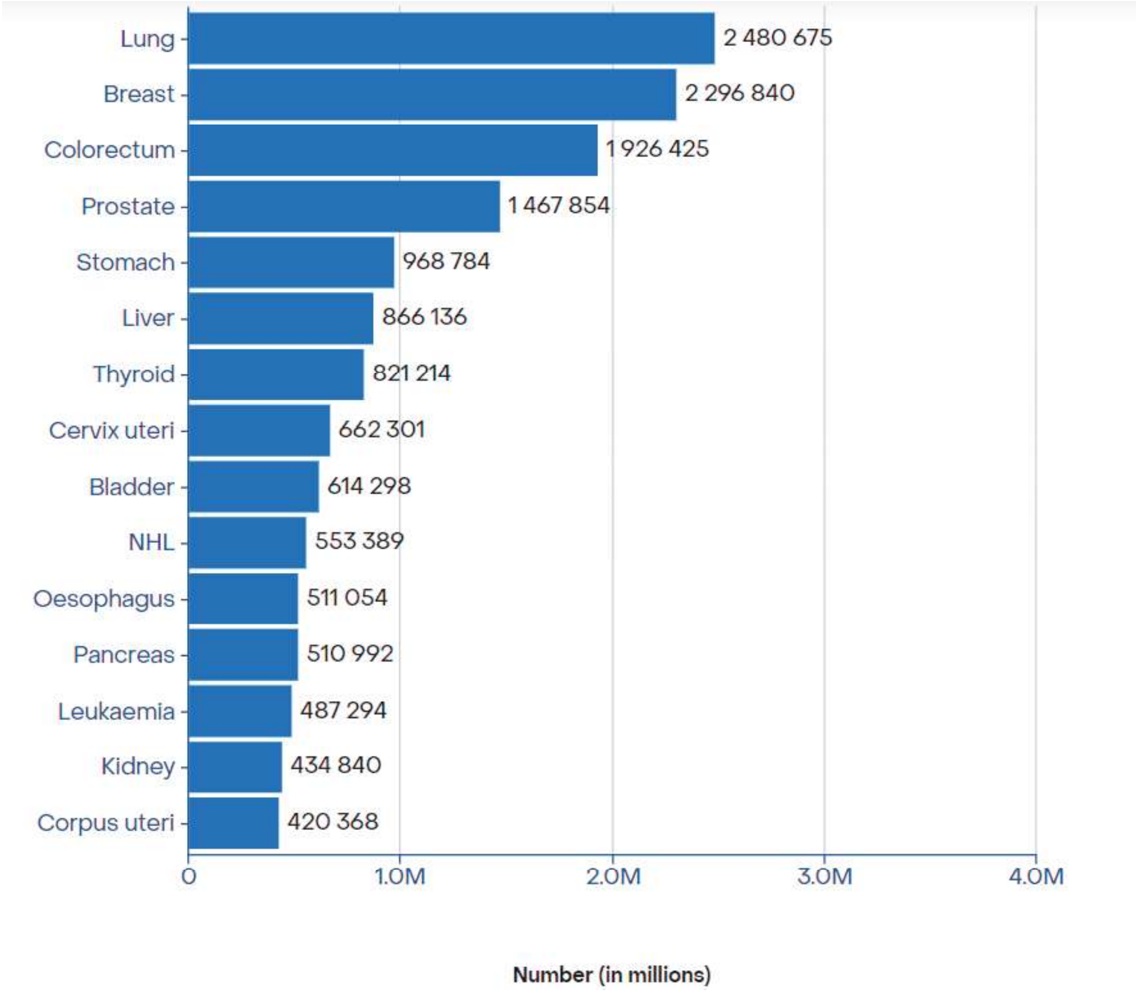


Figure 1. Recorded incidents of different types of cancer worldwide [4]

In the quest for more effective diagnostic tools, artificial intelligence (AI) has emerged as a game-changing technology. Recent studies showed that AI algorithms could outperform human experts in certain medical image diagnosis tasks, including lung diseases [5]. Among the various AI approaches, fuzzy logic stands out for its ability to handle uncertainty and imprecision - characteristics inherent in medical diagnosis. Fuzzy logic is a mathematical logic that attempts to solve problems with explicit, imprecise, or approximate reasoning [6]. It is based on fuzzy set theory, which is an extension of classical set theory [7]. It provides a way to obtain a precise result based on uncertain, ambiguous, imprecise, noisy, or incomplete input information [8] and has gained great importance in disease prediction due to its ability to handle uncertainties and ambiguity in medical data. Fuzzy logic has been widely applied in medical diagnosis, particularly in lung cancer and related conditions, demonstrating its efficacy in improving the diagnostic efficiency of tumor markers [9], enhancing the detection of disease progression in small cell lung cancer patients [10], and even extending to the diagnosis of other respiratory diseases such as pneumonia through expert systems [11]. Overall, fuzzy logic, with its capacity to mimic human reasoning and handle ambiguous data, offers a promising avenue for enhancing lung cancer detection and prediction [12].

Compared with other AI tools, such as Neural Networks and Support Vector Machines (see table 1), fuzzy logic offers unique advantages in handling the uncertainties inherent in medical diagnosis. Its ability to provide interpretable results is particularly valuable in clinical settings where transparency in decision-making is crucial [13].

The challenge of diagnosing lung cancer nodules, coupled with the high mortality rate of lung cancer, underscores the significance of using advanced methods like fuzzy logic [14]. Fuzzy logic offers a problem-solving approach that relies on logical rules and if-then statements, incorporating human experience and expertise [16]. This makes it particularly suited for medical applications where expert knowledge plays a crucial role [9].

Table 1. Comparison of Fuzzy Logic Method with other AI tools

Technique	Strengths	Limitations
Fuzzy Logic	• Handles uncertainty well • Mimics human reasoning • Interpretable results	• Can be complex for large rule sets • Requires expert knowledge for initial setup
Neural Networks	• Powerful pattern recognition • Can handle large datasets • Adaptive learning	• "Black box" nature • Requires large training datasets • Less interpretable
Support Vector Machines	• Effective in high-dimensional spaces • Versatile through kernel trick	• Less intuitive • Can be computationally intensive

Taking into consideration the aforementioned, studies on fuzzy logic are important in diagnosing lung cancer. After a preliminary literature review, the authors identified research papers that deal with the diagnosis of lung cancer using fuzzy logic methods; systematic reviews that deal with disease (general) diagnosis or healthcare with fuzzy logic methods, or, modern approaches used in the detection of lung cancer [16][17][18]. However, no systematic review study has been found that brings together studies on fuzzy logic methods in the diagnosis of lung cancer. In this way, this study constitutes an original one, which examines scholarly work on the specific topic and provides an overview of the literature.

The primary aim of this study is to review the literature on lung cancer diagnosis using fuzzy logic methods. Through this systematic review, we seek to accomplish several objectives:

- To provide researchers and field experts with a comprehensive overview of relevant literature in this domain.
- To present research results, findings, and recommendations in a clear and accessible manner.
- To offer a valuable resource for both researchers and practitioners involved in the diagnosis and treatment of lung cancer.

By achieving these goals, we aim to facilitate advancements in the application of fuzzy logic to lung cancer diagnosis and treatment.

The primary aim of this study is to comprehensively review the literature on lung cancer diagnosis using fuzzy logic methods. Specifically, this systematic review seeks to achieve the following objectives:

- To provide researchers and field experts with a concise overview of relevant literature in the application of fuzzy logic to lung cancer diagnosis.
- To present research results, findings, and recommendations in an accessible manner, serving as a valuable resource for both researchers and practitioners.
- To synthesize information on fuzzy logic methods used in the diagnosis and treatment of lung cancer, offering insights into current practices and future directions.

By collecting and analyzing fuzzy logic methods from various sources, this review aims to offer a broad perspective on the existing literature in this field. Ultimately, it seeks to facilitate advancements in lung cancer diagnosis by consolidating current knowledge and identifying areas for future research.

In this context, the research question of the study is: What is the role of fuzzy logic methods in early detection and prediction of lung cancer, and which fuzzy logic methods are used in this context?

Within the framework of research question, findings from various studies were synthesized, and the effectiveness and potential limitations of fuzzy logic applications in the context of lung cancer prediction were evaluated. Through the synthesis of existing knowledge, the authors aimed at identifying trends, difficulties, and future directions in the use of fuzzy logic methods to improve the accuracy and credibility of lung cancer prediction models. By exploring the intricacies of fuzzy logic-based lung cancer prediction, this review contributes to the ongoing discourse on the use of computational intelligence in healthcare. By critically examining the existing literature, the goal is to provide information that can guide future research efforts and, as a result, stimulate advances in the field of lung cancer prediction and contribute to improved patient outcomes.

The structure of the study is as followed: Section 2 presents a short review of related studies on soft computing methods in medical diagnosis and their critical analysis. Section 3 includes the methodology part for this systematic review. Section 4 presents the results of the research. Finally, section 5 concludes this study with conclusions, future research and limitations.

2. Literature Review

The application of soft computing methods in medical diagnosis, particularly in lung cancer prediction and diagnosis, has been a subject of significant research interest. This section provides a critical analysis of key studies in this field, highlighting the strengths and weaknesses of fuzzy logic methods compared to other soft computing approaches such as neural networks and genetic algorithms.

Fuzzy logic, a cornerstone of soft computing, has demonstrated its efficacy in lung cancer diagnosis. Schneider et al. [9] showed that fuzzy logic-based tumor-marker profiles could improve sensitivity in lung cancer diagnosis from 70% to 90%. This highlights fuzzy logic's ability to handle the imprecision inherent in medical data effectively.

[16] conducted a comprehensive review of fuzzy logic methods in disease diagnosis. While not specific to lung cancer, their study underscored fuzzy logic's strength in managing uncertainty in medical data, a crucial factor in cancer diagnosis.

The potential of neural networks, another key soft computing technique, in lung cancer diagnosis was evident in the meta-analysis by Liu et al. [19]. They found that AI methods, including neural networks, demonstrated high diagnostic accuracy in lung cancer, with a pooled sensitivity of 0.87 and specificity of 0.83.

Genetic algorithms, while less prominently used in direct lung cancer diagnosis, have shown potential in optimizing other soft computing methods. Daliri [20] presented a hybrid system combining genetic algorithms with fuzzy extreme learning machines for lung cancer diagnosis. This study demonstrated how genetic algorithms could optimize the parameters of fuzzy systems, enhancing overall performance.

The power of combining soft computing methods was evident in the study by Lin and Yang [21]. They introduced a Fusion-Based Convolutional Fuzzy Neural Network (F-CFNN) for lung cancer classification, achieving 97% accuracy on a large dataset of 22,489 CT images. This hybrid approach leveraged the pattern recognition strengths of neural networks and the interpretability of fuzzy logic, showcasing the synergistic potential of soft computing techniques.

Thomas et al. [22], in their systematic review, further supported the efficacy of fuzzy models in medical diagnosis, including cancer detection. Their findings reinforced the consistent high accuracy of fuzzy logic across various medical diagnostic applications.

While not specific to lung cancer, studies by Wagner et al. [23] and Jan et al. [24] highlighted the broader application of soft computing methods in oncology, including surgical decision support and early diagnosis of other cancers like pancreatic cancer. These studies underscore the versatility and potential of soft computing techniques in the broader field of cancer diagnosis and treatment.

The aforementioned studies can provide a comprehensive overview of the current state of soft computing methods in lung cancer diagnosis. Thus, Table 2 presents a comparative analysis of fuzzy logic, neural networks, genetic algorithms, and hybrid approaches, highlighting their respective strengths, limitations, and key studies in the field.

Table 2. Comparison of Soft Computing Methods in Lung Cancer Diagnosis

Method	**Strengths**	**Limitations**	**Key Studies**
Fuzzy Logic	• Handles uncertainty and imprecision well • Highly interpretable results • Improved sensitivity in diagnosis • Can incorporate expert knowledge	• May require complex rule sets for nuanced problems • Performance depends on quality of fuzzy rule set design	[9] [16] [22]
Neural Networks	• Excellent pattern recognition • High accuracy in image-based diagnosis • Can handle large, complex datasets	• "Black box" nature limits interpretability • Requires large datasets for training• Prone to overfitting	[19]
Genetic Algorithms	• Effective for parameter optimization • Can improve performance of other soft computing methods • Good at finding global optima	• Computationally intensive • May converge to local optima • Requires careful parameter tuning	[20]

Method	Strengths	Limitations	Key Studies
Hybrid Methods (e.g., Fuzzy Neural Networks)	• Combines strengths of multiple soft computing approaches • High accuracy while maintaining some interpretability • Can overcome limitations of individual methods	• Complex model design and implementation • May require significant computational resources • Potential difficulty in determining optimal hybridization	[21]

Literature review revealed the need for a systematic review and the focus on fuzzy logic method within the framework of lung diagnosis:

- While Ahmadi et al. [16] and Thomas et al. [22] provided broad reviews of fuzzy logic in medical diagnosis, there is a lack of comprehensive reviews specifically focusing on fuzzy logic applications in lung cancer diagnosis.
- Lung cancer diagnosis inherently involves dealing with uncertain and imprecise data. Fuzzy logic's strength in handling such uncertainty, as demonstrated by Schneider et al. [9], makes it particularly suitable for this domain.
- Unlike "black box" methods such as some neural network approaches, fuzzy logic provides interpretable results. This interpretability is crucial in medical applications where understanding the reasoning behind a diagnosis is essential for clinician trust and patient communication.
- Fuzzy logic allows for the direct incorporation of expert knowledge into the system. This is particularly valuable in the medical field where expert opinions play a significant role alongside data-driven insights.
- While some AI methods require large datasets for training, fuzzy logic has shown effectiveness even with smaller datasets. This is advantageous in medical research where large, standardized datasets may not always be available.
- The success of hybrid methods like the one used by Lin et al. [21] suggests that fuzzy logic can be effectively combined with other soft computing techniques to leverage the strengths of multiple approaches.
- The broader success of fuzzy logic in various medical diagnostic applications, as shown by Thomas et al. [22] indicates its potential for further development in lung cancer diagnosis specifically.
- Lung cancer diagnosis often involves complex, interrelated factors. Fuzzy logic's ability to handle complex rule sets makes it well-suited to capture these intricate relationships.

This review aims to address these gaps by providing a comprehensive analysis of fuzzy logic methods in lung cancer prediction and diagnosis, highlighting their unique strengths and potential for integration with other soft computing methods. The novelty of fuzzy logic in this domain lies in its ability to handle medical uncertainties in an interpretable manner, a crucial factor in clinical decision-making processes.

3. Methodology

This research was prepared according to the systematic review method and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) technique used as a guide for systematic review. The PRISMA flowchart is an essential tool for conducting systematic reviews and meta-analyses. It provides a structured approach for documenting the selection process of studies, ensuring transparency and reproducibility in the research. The aim is to improve the accuracy and precision of such inquiries by giving originators a checklist to guarantee a through disclosure. The use of the PRISMA flowchart is recommended in various fields, including psychology, medicine, biology, and education [25]. Adherence to PRISMA rules is associated with improved quality of detailing in efficient audits [26]. An effective review strategy was applied to this research, and the articles included in the research were recovered from the database using the PRISMA strategy. The research utilized popular databases, including Scopus, Web of Science, and PubMed, renowned for their extensive coverage and relevance in the healthcare domain. The keywords used to search the databases are given in Table 1 and the inclusion and exclusion criteria are given in Table 2.

In the inclusion process of the studies included in the study, firstly, keywords suitable for the research question were searched from the databases related to the studies, and after the studies were eliminated in accordance with the inclusion and exclusion criteria, the remaining studies were exported and the titles and abstracts of the studies were first examined. Studies whose titles and abstracts were not appropriate for the research questions were also eliminated and the full text of the remaining studies was examined. From the studies whose full text was examined, the studies that were appropriate for the research question were selected and finally, the studies to be included in the systematic review process were decided.

3.1. Search keywords for databases

The queries for the keywords used in the database search are given in Table 3. The search queries were meticulously crafted by experimenting with different combinations to optimize search results based on the databases' structures.

Table 3. Search keywords

Scopus	Web of Science	Pubmed	IEEE Xplore
(TITLE-ABS-KEY (fuzzy) AND TITLE-ABS-KEY (“lung cancer” OR “lung cancer prediction” OR “lung cancer predicting”) AND TITLE-ABS-KEY (predicting OR prediction))	fuzzy (Title) and (“lung cancer” OR “lung cancer prediction” OR “lung cancer predicting”) (Title) or fuzzy (Abstract) and (“lung cancer” OR “lung cancer prediction” OR “lung cancer predicting”) (Abstract) and predicting or prediction (Abstract)	TI fuzzy AND TI (“lung cancer” or “lung cancer predicting” or “lung cancer prediction”) OR AB fuzzy AND AB (“lung cancer” or “lung cancer predicting” or “lung cancer prediction”) AND (predicting or prediction)	(“Document Title”: fuzzy) AND (“Document Title”: “lung cancer” OR “Document Title”: “lung cancer prediction” or “lung cancer predicting”) OR (“Abstract”: fuzzy) AND (“Abstract”: “lung cancer” OR “Abstract”: “lung cancer prediction” or “lung cancer predicting”) AND (“All Metadata”: predicting or prediction)

3.2. Refinement of initial results (inclusion and exclusion criteria)

The articles retrieved from the database searches underwent screening based on specific inclusion and exclusion criteria to ensure relevance to the study. To work with current data, articles between 2019 and 2024 and in English were included in the study. On the other hand, papers, books and book chapters and review type publications were not included in the study. Articles from the PubMed database were not included in the study because they were considerably older than 2019. Similarly, articles accessed from the Web of Science database were also eliminated based on date and publication type. Many of the articles included in the study were obtained from the Scopus database.

Table 4. Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Language: English, Article	Review articles, Book, Book section, Dublicate studies, Thesis, Conference Paper, Retracted Publication, Conference Review, Article in press

3.3. Validity and Reliability

The validity of the research method was confirmed through a thorough keyword selection process and database search during the literature review. In addition, filtering criteria appropriate to the objectives of the study were created, and the validity of the research was strengthened by analyzing the articles that met these requirements. The research's reliability was supported by its reproducibility, prior preliminary review, and consistent results from independent researchers, demonstrating its credibility.

4. Results and Discussion

This section presents the findings resulting from the systematic review conducted to address the research questions. The articles retrieved from the search were screened based on predefined inclusion and exclusion criteria and specific database filtering methods. This information is given in the Identity section of the PRISMA flow chat in Figure 2. The abstracts of the remaining 86 articles were examined, and the articles to be read in full text were decided, and the number of articles was reduced to 55. This information is shown in the Screening section. After reviewing the full text of 55 articles, the final selection for inclusion in the study was made based on the predetermined criteria. Finally, this information is given in the Included section, and the full text of 49 articles was examined in the study.

4.1. Observations on Datasets

The studies in the table also vary in terms of datasets and sample sizes used for lung cancer diagnosis. While some studies conduct comprehensive analyses on large datasets, others focus on smaller and more specific sample groups. This variability influences the scope of each study and the generalizability of the findings. For instance, some studies leverage CT scans, X-ray images, and gene expression data to evaluate the effectiveness of various diagnostic and classification methods. Additionally, the findings from these studies highlight the continuous development of methods used in lung cancer diagnosis. Each study is designed to address a specific problem or optimize a particular approach, contributing to the overall knowledge base in lung cancer diagnosis. It is observed that recent studies employ more advanced methods and sophisticated models compared to earlier ones, leading to higher accuracy rates and more reliable results in lung cancer diagnosis.

In this part of the research, information about the data sets of the studies included in the systematic review is given.

LIDC-IDRI (Lung Image Database Consortium and Image Database Resource Initiative) is a consortium dataset for early detection of lung nodules and cancer diagnosis. It focuses on the identification of lung nodules of different sizes and types using CT scans. Since this dataset contains various features such as the size, shape, and density of lung nodules, it allows fuzzy logic algorithms to manage these uncertainties. The ambiguous boundaries of the nodules provide a suitable space for fuzzy logic to better define the boundaries [27].

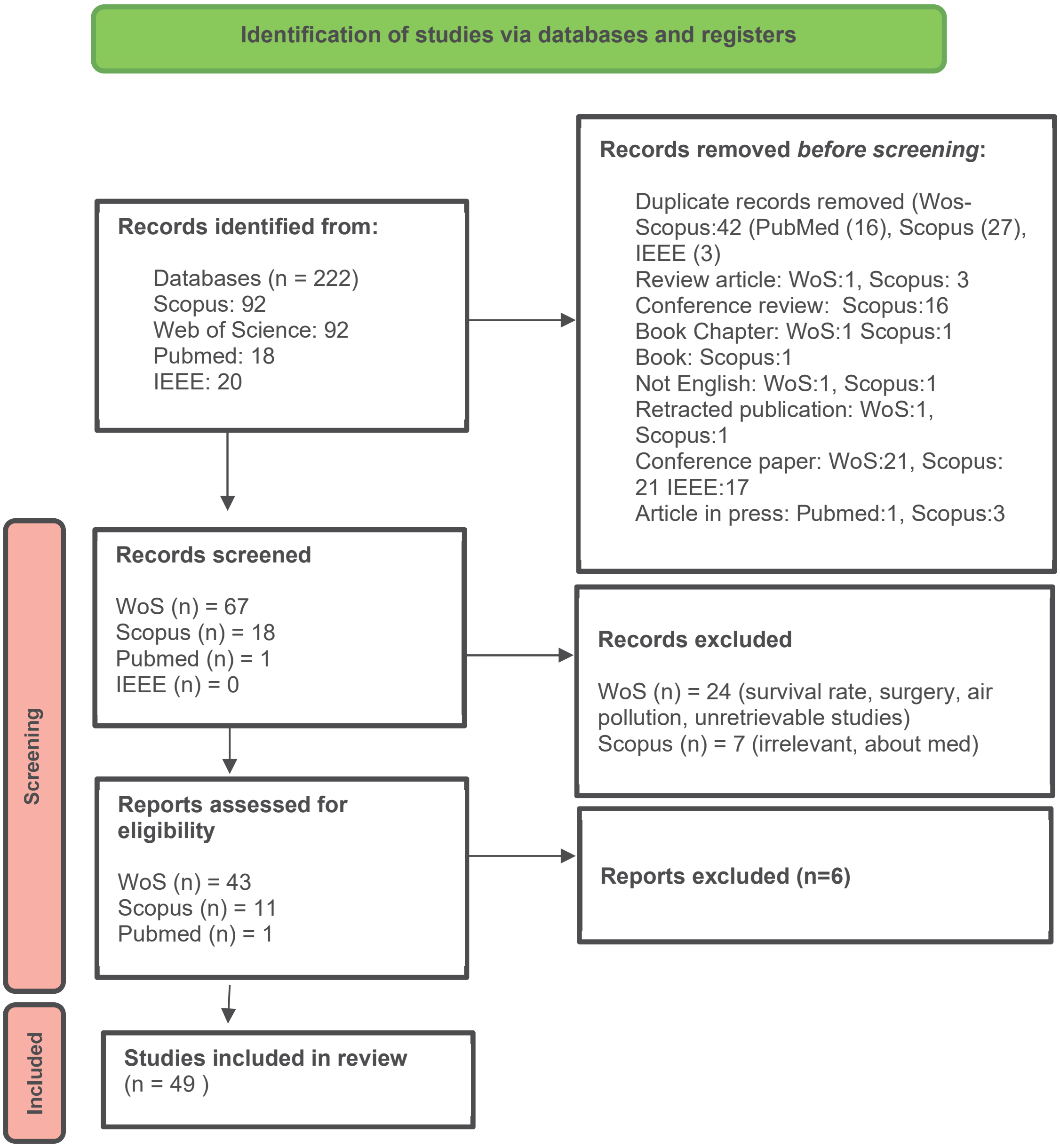


Figure 2. PRISMA flow diagram [72]

SPIE-AAPM Lung CT Challenge Dataset and LIDC-IDRI Lung Imaging Research Dataset; Organized by the American Association for Medical Physics (AAPM) and SPIE, this dataset aims to use computed tomography (CT) images for lung cancer detection [28]. This dataset helps fuzzy logic algorithms to improve classification and prediction accuracy because of the uncertainties between different cancer stages, nodule sizes and cancer types in a wide variety of CT images. The dataset also provides an ideal testbed for diagnosing cancer at different stages.

COVID-19 Detection X-Ray Dataset and Lung Cancer Dataset; The purpose of this dataset shared on Kaggle is to investigate the differences between COVID-19 detection and lung cancer. This dataset, which includes X-ray images, contributes to the diagnosis of various lung diseases. X-ray images can have ambiguities due to low resolution and noise. These ambiguities can be processed with fuzzy logic algorithms to achieve better classification results. Moreover, similar symptoms of diseases and similarities in images require fuzzy logic to reduce uncertainties in decision making [27].

Lung Cancer Gene Expression Dataset: This dataset, obtained from the UCI Machine Learning Repository, is used to analyze gene expression profiles of lung cancer. It is used in lung cancer diagnosis based on genetic analysis. Gene expression data includes analysis of genes expressed at different levels in each individual patient. This data set shows a large variability that needs to be managed with fuzzy logic, as genetic data contains uncertainties. Given that small changes in gene expression data can have major clinical consequences, fuzzy logic allows for precise handling of these variables [29].

IQ-OTH/NCCD Dataset; This dataset contains CT images and clinical data for the diagnosis of lung cancer and is specifically used for the accurate detection of nodules. Managing uncertainties in CT images can be supported by fuzzy logic algorithms for accurate classification of nodules [30].

TCIA is an open access archive that collects and shares various imaging data related to cancer. It aims to provide researchers with large-scale imaging data and contribute to innovative studies based on medical image analysis [31]. The complexity and diversity of images allows for managing uncertainties and making more precise classifications with fuzzy logic.

Open datasets published by the global burden of disease; This open dataset published by the Global Burden of Disease (GBD) study is a comprehensive database on the burden of disease and health problems worldwide. It includes global health statistics for various diseases, such as lung cancer, and is provided to researchers for use in public health analysis. Fuzzy logic techniques can be applied to manage uncertainties and incompleteness in the data for public health analyses. The uncertain and complex nature of the data makes fuzzy logic approaches valuable in health services planning and policy making [32].

Random Sample of NIH Chest X-ray Dataset; created by the NIH to diagnose lung diseases. X-ray images are used to classify lung cancer as well as other respiratory diseases. Fuzzy logic techniques can be used to correctly classify ambiguous areas in X-ray images. Factors such as image quality and noise can be managed with fuzzy logic algorithms [33].

CIA Datasets Cancer Imaging Archive; are a set of datasets containing imaging data for cancer. These datasets are based on various imaging modalities (CT, MRI, PET scans, etc.) used in cancer diagnosis and treatment. These datasets contain medical imaging data on many types of cancer, especially lung cancer. CIA datasets are suitable for modeling uncertainties and variability in imaging data. In imaging diseases such as lung cancer, fuzzy logic allows for more accurate diagnoses and classifications by managing uncertainties [34].

The ILD (Interstitial Lung Disease) dataset contains CT scan images for interstitial lung diseases. This dataset is specifically used to analyze and classify changes in lung tissues. It is suitable for fuzzy logic algorithms in terms of fuzzy boundaries and uncertainty management. It allows accurate classification of interstitial lung diseases, which may have similar image characteristics to lung cancer [28].

The dataset provided by the **Lung Cancer Alliance** is a database for the diagnosis and treatment of lung cancer. Fuzzy logic can be used to model uncertainties in genetic and clinical data related to the disease. This dataset can be analyzed with fuzzy logic methods, especially to distinguish between different types of cancer. On the other hand, the limited number of rare cases may make it difficult to generalize the modeling [35].

The Kentridge Biomedical Repository is a database of gene expression data used in cancer research. This dataset was developed to be used in the diagnosis and classification of various types of cancer, including lung cancer. The genetic data it contains enables the application of classification algorithms in cancer research. It can be said that this dataset, which contains high-dimensional gene expression data, is very suitable for fuzzy logic models. It is especially used to manage uncertainties in the relationships between genes and to eliminate unnecessary genes. Fuzzy logic-based approaches improve the classification accuracy of small differences in gene expression. The large size of the data and the high dimensionality of genetic data can increase processing and storage costs [36].

Microarray gene expression datasets are used to analyze gene expression data on a large scale and diagnose diseases, especially cancer. These datasets are also frequently used in lung cancer diagnosis and enable disease identification at the gene level. These data sets are collected by various laboratories and research centers. Microarray datasets contain large variability and uncertainty in gene expression data. This makes fuzzy logic models particularly suitable for classifying genetic variations and small changes in expression levels. Fuzzy logic plays a critical role in managing these uncertainties and ensuring accurate classification [37].

The UCI Machine Learning Repository is a large archive of datasets collected for testing with various machine learning algorithms. This dataset for lung cancer diagnosis contains 32 samples and 3 different pathological types of lung cancer. There are 56 features for each sample. Although this dataset contains a limited number of samples, it has enabled fuzzy logic algorithms to achieve high success even with small samples. Fuzzy logic rules can be effective in classifying the fuzzy boundaries of different cancer types. However, since the dataset has a small sample size, its generalization ability is limited. Missing data points can make the analysis more complex, so the missing features were filled by averaging [20]. These datasets are united by common problems such as large dataset sizes, which increase storage and processing demands, and variability in data quality, which can affect model accuracy. Furthermore, the scarcity of data on rare cancer cases is a barrier to generalizing the findings. Nevertheless, these datasets are vital for early detection of lung cancer, improving patient outcomes and informing public health strategies. Furthermore, integrating fuzzy logic into the analysis of these datasets can help remove inherent uncertainties in medical data, improve classification accuracy and support the development of clinical decision support systems. This can contribute to better diagnostic processes, more personalized treatments, and improved public health policies.

4.2. Findings on Employed Methods

In this part, information is given about the articles that have been included in the research after the systematic review. Appendix-1 lists the years, aims, results, and fuzzy logic methods of the research included in the systematic review. In addition, details regarding the diagnosis of lung cancer are also given according to the studies performed. Therefore, researchers who review this article and want to examine lung cancer diagnosis using the fuzzy logic method can get an idea by looking at this table.

The systematic review presented in the table encompasses 49 different studies focusing on the integration of fuzzy logic and deep learning approaches in the diagnosis and treatment of lung cancer. Each study addresses unique methods developed for specific purposes and their outcomes, contributing to the existing approaches in lung cancer diagnosis and treatment. A significant portion of the studies in the table utilize fuzzy logic systems and hybrid methods to manage uncertainties and improve accuracy in complex problems like lung cancer diagnosis and classification. Overall, the studies emphasize the effectiveness of fuzzy logic systems, particularly in handling data with inherent uncertainties. Complex medical problems, such as lung cancer diagnosis and classification, often involve uncertainties and variabilities that traditional algorithms struggle to manage. Therefore, the frequent use of fuzzy logic-based approaches in these studies highlights the importance of such methods in dealing with uncertainties in medical data. Techniques such as fuzzy rule-based systems, fuzzy clustering methods, and fuzzy inference systems (FIS) are commonly employed in these studies to manage uncertainties in medical imaging, disease classification, and decision-making processes.

The table also reveals a widespread use of hybrid models. These hybrid approaches integrate fuzzy logic systems with deep learning, optimization techniques, and other artificial intelligence methods to achieve higher accuracy and more robust outcomes. For example, in some studies, fuzzy logic systems are combined with convolutional neural networks or other clustering methods to improve the accuracy of lung cancer diagnosis and classification. These hybrid approaches play a crucial role in improving the precision of complex medical diagnoses like lung cancer, where traditional methods may fall short.

Built upon the overview of studies presented in Table 5, the following table provides a comprehensive comparison of these fuzzy logic methods, offering researchers and clinicians a clearer understanding of their relative merits and challenges. By examining the strengths and limitations of each approach, we can gain valuable insights into their effectiveness and potential applications in clinical settings.

Tablo 5. Comparative analysis of fuzzy logic methods used in lung cancer diagnosis

Method	Pros	Cons	Example Studies
Fuzzy Rule-Based Systems (FRBS)	- Highly interpretable - Can incorporate expert knowledge - Flexible and adaptable to various input types	- Limited performance compared to complex ML models - Time-consuming rule base design - Struggles with high-dimensional data	[26], [38]

Method	Pros	Cons	Example Studies
Fuzzy C-Means Clustering (FCM)	- Effective for image segmentation - Allows partial membership - Faster than hierarchical clustering	- Sensitive to initial conditions- Requires pre-specified cluster number - Struggles with imbalanced datasets	[31], [39]
Fuzzy Inference Systems (FIS)	- Handles complex, non-linear relationships - Combines interpretability with adaptivity - Effective for classification and regression	- Complex to design and tune - Computationally intensive for large datasets - Decreasing interpretability with complexity	[32], [40]
Adaptive Neuro-Fuzzy Inference Systems (ANFIS)	- Combines neural network learning with fuzzy interpretability - Automatically adjusts membership functions - Effective for complex, non-linear problems	- Computationally intensive- Risk of overfitting- Complex models can be hard to interpret	[41], [42]
Hybrid Fuzzy-Based Approaches	- Leverages strengths of multiple techniques - Higher accuracy than individual fuzzy methods - Handles diverse data types and complex patterns	- Increased model complexity - Extensive hyperparameter tuning - Risk of overfitting with limited data	[21], [20]
Fuzzy-Enhanced Deep Learning Models	- Combines deep learning with fuzzy uncertainty handling - State-of-the-art performance on complex tasks - Handles large-scale, high-dimensional data	- Requires significant computational resources - Less interpretable than traditional fuzzy systems - May require large datasets for optimal performance	[28], [43]

Having examined the general characteristics, strengths, and limitations of various fuzzy logic methods applied in lung cancer diagnosis, the study provides then specific implementations of these methods that have demonstrated superior performance in our reviewed. These top-performing approaches, which include specialized applications of fuzzy-enhanced deep learning models, hybrid fuzzy-based approaches, and advanced fuzzy inference systems, offer unique advantages in addressing the challenges of lung cancer diagnosis. This analysis gives insights into their practical implications and clinical relevance, revealing their potential impact on real-world lung cancer diagnostics (see table 6).

Tablo 6. Comparative Analysis of Top-Performing Fuzzy Logic Approaches in Lung Cancer Diagnosis

Approach	Performance Metrics	Key Strengths	Practical Implications	Clinical Relevance
Fusion-Based Convolutional Fuzzy Neural Network (F-CFNN) [21]	Accuracy: 97.2% Precision: 96.8% Recall: 97.5% F1 Score: 97.1%	- High accuracy- Handles complex image data - Combines strengths of CNN and fuzzy logic	- Scalable to large datasets - Requires significant computational resources - Potential for real-time application with optimized hardware	- Excellent for early detection - High accuracy in distinguishing malignant from benign nodules - Adaptable to diverse patient demographics
Fuzzy K-Nearest Neighbor (FKNN) with Enhanced Manta Ray Foraging Optimization [38]	Accuracy: 95.8% Sensitivity: 94.3% Specificity: 97.2% AUC: 0.982	- Robust to noise in data - Effective handling of uncertainty - Improved optimization through EMRFO	- Moderate computational requirements - Easily interpretable results	- Suitable for risk stratification - Effective in cases with ambiguous imaging results - Adaptable to different types of medical data

Approach	Performance Metrics	Key Strengths	Practical Implications	Clinical Relevance
			- Potential for integration with existing medical systems	
Deep Fuzzy SegNet [43]	Accuracy: 98.6% Dice Coefficient: 0.945 Jaccard Index: 0.896	- Excellent segmentation performance - Handles complex lung structures - Integrates deep learning with fuzzy logic	- Requires specialized hardware for optimal performance - Potential for automated analysis in clinical workflows - Scalable to large-scale screening programs	- High precision in nodule detection - Assists in treatment planning through accurate segmentation - Potential for tracking tumor changes over time
Fuzzy Soft Expert System [44]	Accuracy: 93.5% Sensitivity: 92.1% Specificity: 94.8%	- Incorporates expert knowledge - Handles uncertainty in clinical data - Highly interpretable results	- Low computational requirements - Easy integration with existing clinical decision support systems - Adaptable to new expert knowledge	- Effective for initial risk assessment - Aids in personalized treatment planning - Suitable for diverse clinical settings, including resource-limited areas

The F-CFNN and Deep Fuzzy SegNet approaches demonstrate the highest overall accuracy, particularly in image-based diagnosis. The FKNN with EMRFO shows excellent balance between sensitivity and specificity, making it robust for general screening purposes. The Fuzzy Soft Expert System, while having slightly lower accuracy, offers high interpretability which is crucial in clinical settings.

- Strengths and Limitations

The reviewed fuzzy logic approaches demonstrate diverse strengths and limitations in lung cancer diagnosis. F-CFNN [21] and Deep Fuzzy SegNet [43] stand out for their exceptional ability to process and analyze complex image data, making them particularly effective for interpreting medical imaging results. However, these sophisticated models come with the drawback of requiring substantial computational resources, which may limit their accessibility in some clinical settings. In contrast, FKNN with EMRFO [38] offers a compelling middle ground, striking a balance between high performance and interpretability. This balance makes it a versatile option suitable for a wide range of clinical applications, from initial screening to more detailed diagnostic processes. The Fuzzy Soft Expert System [44], while perhaps less adept at handling complex image analysis tasks, shines in its capacity to incorporate expert knowledge directly into the diagnostic process. Its high interpretability is a significant advantage, particularly in scenarios where clear explanation of the diagnostic reasoning is crucial for patient care and clinical decision-making.

- Practical Implications

The practical implementation of these fuzzy logic approaches varies considerably based on their computational requirements and the clinical context. F-CFNN [21] and Deep Fuzzy SegNet [43], with their advanced capabilities in image analysis, are ideally suited for deployment in large hospitals or specialized imaging centers equipped with robust computational infrastructure. These settings can leverage the full potential of these models to enhance diagnostic accuracy in complex cases. FKNN with EMRFO [38], thanks to its more moderate computational demands, presents an attractive option for integration into existing clinical workflows across a broader range of healthcare facilities. Its balance of performance and resource requirements makes it a practical choice for many medical institutions looking to enhance their diagnostic capabilities without overhauling their entire technological infrastructure. The Fuzzy Soft Expert System [44] with its minimal computational needs and high interpretability, emerges as an excellent candidate for use in primary care settings or as a first-line screening tool. Its ability to provide clear, understandable results makes it particularly valuable in contexts where immediate interpretation and explanation of results to patients is necessary.

- Clinical Relevance

From a clinical perspective, all the examined fuzzy logic methods demonstrate significant promise in the critical area of early lung cancer detection, a factor that is paramount in improving overall patient outcomes. The advanced image analysis capabilities of F-CFNN [21] and Deep Fuzzy SegNet [43] render them particularly effective in the crucial task of distinguishing between malignant and benign nodules. This high level of discrimination can play a vital role in reducing the number of unnecessary biopsies, thereby minimizing patient stress and healthcare costs. FKNN with EMRFO [38] and the Fuzzy Soft Expert System [44] offer a different but equally important clinical advantage: their flexibility in handling various types of clinical data. This adaptability makes them valuable across a wide spectrum of clinical scenarios, from initial patient screening to ongoing monitoring of high-risk individuals. Their ability to integrate diverse data types allows for a more comprehensive approach to lung cancer diagnosis, potentially capturing subtle indicators that might be missed by more narrowly focused diagnostic tools.

The field of fuzzy logic in lung cancer diagnosis presents several promising avenues for future research and development. One key direction is the integration of these fuzzy logic approaches with other AI technologies, such as natural language processing for analyzing clinical notes, which could significantly enhance their diagnostic capabilities [27]. This integration could allow for a more comprehensive analysis of patient data, incorporating both structured and unstructured information. Additionally, the development of lightweight versions of more complex models like F-CFNN [21] and Deep Fuzzy SegNet [43] could broaden their applicability in resource-limited settings, making advanced diagnostic tools more accessible to a wider range of healthcare facilities. This democratization of technology could have far-reaching implications for early lung cancer detection globally. Furthermore, continued research into improving the interpretability of complex fuzzy-deep learning hybrid models is crucial. Enhancing the explainability of these sophisticated systems could increase their acceptance in clinical practice [13], addressing one of the key challenges in the adoption of AI in healthcare.

In conclusion, this systematic review underscores the significant role that fuzzy logic and artificial intelligence-based methods play in lung cancer diagnosis and classification. The variety of techniques and approaches used in these studies demonstrates the unique advantages and challenges associated with each method, while the general trend points towards the development of more integrated and hybrid solutions. These hybrid approaches are proving to be effective tools in managing uncertainties and achieving more accurate results in the diagnosis of complex diseases like lung cancer. As these models become more transparent and their decision-making processes more understandable to clinicians, their integration into routine clinical workflows is likely to accelerate, potentially leading to improved patient outcomes through more accurate and timely diagnoses. Future research is likely to focus on further developing these hybrid methods and expanding their application areas. In this context, the studies presented in this review serve as a valuable resource highlighting the potential of fuzzy logic and artificial intelligence techniques in lung cancer diagnosis and treatment. The continued evolution of these technologies promises to significantly enhance our ability to detect, diagnose, and ultimately improve outcomes for patients with lung cancer.

5. Conclusion and Future Research

In conclusion, this systematic review has provided a comprehensive overview of the role of fuzzy logic methods in the diagnosis and prediction of lung cancer. Fuzzy logic has proven to be a valuable tool in managing the uncertainties inherent in medical data, making it particularly effective in the complex and often ambiguous process of diagnosing lung cancer. The studies reviewed demonstrated the effectiveness of various fuzzy logic methods, including Fuzzy Rule-Based Systems, Fuzzy Clustering, and Fuzzy Inference Systems, in improving the accuracy and reliability of lung cancer diagnosis. The integration of fuzzy logic with other computational techniques, such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and genetic algorithms, has led to the development of hybrid models that further enhance diagnostic accuracy. These hybrid approaches leverage the strengths of each method, resulting in more robust and reliable diagnostic tools.

5.1. Limitations

Despite the promising results, the review also highlighted several important limitations. A significant concern is the issue of sample size and generalizability. Many studies in this review used relatively small sample sizes, which limits the generalizability of their results. For instance, Khalil et al. [44] used only 45 patients in their study, while Daliri [20] used just 32 samples. Such small sample sizes may not adequately represent the diverse population of lung cancer patients, potentially leading to overfitting and reduced performance when applied to larger, more diverse datasets.

Another notable limitation is the lack of extensive clinical validation. While many studies demonstrated high accuracy in controlled settings, the performance of these fuzzy logic-based models in real-world clinical environments, where data may be noisier and more variable, remains largely untested. This gap between research findings and clinical application needs to be addressed to ensure the practical utility of these methods.

The variability in datasets used across studies poses another challenge. The studies reviewed used a wide variety of datasets, making direct comparisons between methods challenging. Some studies used publicly available datasets, while others used proprietary or locally collected data, further complicating the assessment of generalizability.

There is also a noticeable focus on image-based diagnosis, particularly using CT scans, in the majority of the studies. This leaves a gap in research related to non-image-based diagnostic methods, such as genetic and metabolomic data analysis, which could provide valuable complementary information for lung cancer diagnosis and prediction.

Lastly, while fuzzy logic is generally more interpretable than "black box" machine learning models, some of the more complex hybrid models may still pose interpretability challenges for clinicians. Balancing model complexity and interpretability remains an ongoing challenge in the field.

5.2. Future Research Directions

To address these limitations and further advance the field, several promising avenues for future research emerge. Large-scale clinical validation studies should be a priority. Future research should focus on validating promising fuzzy logic-based models on large, diverse patient populations in clinical settings. For example, a multi-center study could be conducted to test the performance of the Fusion-Based Convolutional Fuzzy Neural Network (F-CFNN) proposed by Lin et al. [21] on a dataset of 10,000+ patients from various demographic backgrounds.

The development of standardized benchmark datasets for lung cancer diagnosis would enable more direct comparisons between different fuzzy logic approaches. Researchers could collaborate with organizations like The Cancer Imaging Archive (TCIA) to create and maintain such datasets, fostering more comparable and reproducible research in the field.

Integration of multi-modal data presents another exciting direction for future studies. Researchers should explore the integration of multiple data types, combining imaging data with genetic, metabolomic, and clinical data. For instance, a hybrid model could be developed that combines the image analysis capabilities of the Deep Fuzzy SegNet [43] with gene expression analysis using Fuzzy Min-Max Neural Networks [45]. This multi-modal approach could provide a more comprehensive and accurate diagnostic tool.

Improving the explainability of fuzzy logic models, especially for complex hybrid systems, should be a priority. Future research could focus on creating visualization tools that explain the decision-making process of fuzzy logic systems in a way that is intuitive for clinicians. This would not only increase the trust in these systems but also potentially provide new insights into the diagnostic process.

Longitudinal studies represent another important area for future research. Most current research focuses on single time-point diagnosis. Future studies should explore the use of fuzzy logic in predicting lung cancer progression over time, potentially integrating with electronic health records for continuous monitoring. This could lead to more personalized and adaptive treatment strategies.

The application of fuzzy logic in personalized medicine for lung cancer treatment is a promising avenue. Research into how fuzzy logic can be applied to personalize treatment plans based on individual patient characteristics and tumor profiles could lead to more effective and tailored therapeutic approaches.

Investigating transfer learning with fuzzy logic models could significantly advance the field. Research into how fuzzy logic models trained on one type of cancer or medical condition can be adapted for lung cancer diagnosis could lead to more robust and generalizable models, potentially addressing the issue of limited dataset sizes in some studies.

Finally, as healthcare moves towards more distributed systems, research into how fuzzy logic models can be optimized for edge computing devices could enable real-time, point-of-care lung cancer risk assessment. This could potentially lead to earlier detection and intervention, particularly in resource-limited settings.

By addressing these research directions, the field can move towards more robust, clinically validated, and widely applicable fuzzy logic-based systems for lung cancer diagnosis and prediction. This could ultimately lead to earlier detection, more personalized treatment plans, and improved patient outcomes, marking a significant advancement in the fight against lung cancer.

References

[1] World Health Organization (2022). Cancer. https://www.who.int/news-room/fact-sheets/detail/cancer.

[2] Aberle, D. R., Adams, A. M., Berg, C. D., Black, W. C., Clapp, J. D., Fagerstrom, R. M., ... & National Lung Screening Trial Research Team. (2011). Reduced lung-cancer mortality with low-dose computed tomographic screening. New England Journal of Medicine, 365(5), 395-409.

[3] Chen, W., Wang, Y., Tian, D., & Yao, Y. (2023). CT lung nodule segmentation: A comparative study of-data preprocessing and deep learning models. IEEE Access, 11, 34925-34931.

[4] Cancer today. (2022). International Agency for Research on Cancer. Retrieved September 26, 2024, from https://gco.iarc.who.int/today/en/dataviz/bars?mode=cancer&key=total&group_populations=1&types=0&sort_by=value1&populations=900&multiple_populations=0&values_position=out&cancers_h=39

[5] Cheng, J., Chen, W., Cao, Y., Xu, Z., Tan, Z., Zhang, X., … & Feng, J. (2020). Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. Nature Communications, 11(1), https://doi.org/10.1038/s41467-020-18685-1

[6] Zadeh, L. A. (2008). Is there a need for fuzzy logic? Information Sciences, 178(13), 2751-2779.

[7] Duricova, L., & Hromada, M. (2016). Fuzzy logic as support for security and safety solution in soft targets. MATEC Web of Conferences, 76, 02034. https://doi.org/10.1051/matecconf/20167602034

[8] Ren, Q., Baron, L., & Balazinski, M. (2011). Type-2 fuzzy modeling for acoustic emission signal in precision manufacturing. Modelling and Simulation in Engineering, 2011, 1-12. https://doi.org/10.1155/2011/696947

[9] Schneider, J., Bitterlich, N., Velcovsky, H., Morr, H., Katz, N., & Eigenbrodt, E. (2002). Fuzzy logic-based tumor-marker profiles improved sensitivity in the diagnosis of lung cancer. International Journal of Clinical Oncology, 7(3), 145-151. https://doi.org/10.1007/s101470200021

[10] Schneider, J., Peltri, G., Bitterlich, N., Philipp, M., Velcovsky, H., Morr, H., … & Eigenbrodt, E. (2003). Fuzzy logic-based tumor marker profiles improved sensitivity of the detection of progression in small-cell lung cancer patients. Clinical and Experimental Medicine, 2(4), 185-191. https://doi.org/10.1007/s102380300005

[11] Arani, L. A., Sadoughi, F., & Langarizadeh, M. (2019). An expert system to diagnose pneumonia using fuzzy logic. Acta Informatica Medica, 27(2), 103. https://doi.org/10.5455/aim.2019.27.103-107

[12] Kumar, A., & Jha, P. (2022). Fuzzy logic applications in healthcare: A review-based study. In Next Generation Communication Networks for Industrial Internet of Things Systems (pp. 1-25).

[13] Chang, T. C. (2023). A fuzzy evaluation approach to determine superiority of deep learning network system in terms of recognition capability: case study of lung cancer imaging. Annals of Operations Research, 1-21.

[14] Black, W. C., Gareen, I. F., Soneji, S. S., et al. (2014). Computed tomography screening for lung cancer: How many false positive screens are acceptable? JAMA, 312(23), 2496-2497.

[15] Rivera, M. M., Olvera-González, E., & Escalante, N. (2023). Upafuzzysystems: A python library for control and simulation with fuzzy inference systems. Machines, 11(5), 572.

[16] Ahmadi, H., Gholamzadeh, M., Shahmoradi, L., Nilashi, M., & Rashvand, P. (2018). Diseases diagnosis using fuzzy logic methods: A systematic and meta-analysis review. Computer Methods and Programs in Biomedicine, 161, 145-172.

[17] Kumar, A., & Jha, P. (2022). Fuzzy logic applications in healthcare: A review-based study. In Next Generation Communication Networks for Industrial Internet of Things Systems (pp. 1-25).

[18] Pandey, S. K., & Bhandari, A. K. (2023). A systematic review of modern approaches in healthcare systems for lung cancer detection and classification. Archives of Computational Methods in Engineering, 30, 4359-4378. https://doi.org/10.1007/s11831-023-09940-x

[19] Liu, M., Wu, J., Wang, N., Zhang, X., Bai, Y., Guo, J., … & Tao, K. (2023). The value of artificial intelligence in the diagnosis of lung cancer: a systematic review and meta-analysis. Plos One, 18(3), e0273445. https://doi.org/10.1371/journal.pone.0273445

[20] Daliri, M. R. (2012). A hybrid automatic system for the diagnosis of lung cancer based on genetic algorithm and fuzzy extreme learning machines. Journal of medical systems, 36, 1001-1005.

[21] Lin, C. J., & Yang, T. Y. (2023). A fusion-based convolutional fuzzy neural network for lung cancer classification. International Journal of Fuzzy Systems, 25(2), 451-467.

[22] Thomas, L. L., Goni, I. and Emeje G. D. (2019). Fuzzy Models Applied to Medical Diagnosis: A Systematic Review. Advances in Networks, 7(2), 45-50. https://doi.org/10.11648/j.net.20190702.15

[23] Wagner, M., Schulze, A., Haselbeck-Köbler, M., Probst, P., Brandenburg, J. M., Kalkum, E., ... & Müller-Stich, B. P. (2022). Artificial intelligence for decision support in surgical oncology-a systematic review. Artificial Intelligence Surgery, 2(3), 159-172.

[24] Jan, Z., El Assadi, F., Abd-Alrazaq, A., & Jithesh, P. V. (2023). Artificial intelligence for the prediction and early diagnosis of pancreatic cancer: Scoping review. Journal of Medical Internet Research, 25, e44248.

[25] Croitoru, D., Huang, Y., Kurdina, A., Chan, A., & Drucker, A. M. (2019). Quality of reporting in systematic reviews published in dermatology journals. British Journal of Dermatology, 182(6), 1469-1476. https://doi.org/10.1111/bjd.18528

[26] Majumder, S., Gautam, N., Basu, A., Sau, A., Geem, Z. W., & Sarkar, R. (2024). MENet: A Mitscherlich function based ensemble of CNN models to classify lung cancer using CT scans. *Plos one*, *19*(3), e0298527.

[27] Kumar, S. J. N., Alam, M. G., Raj, T. M., & Mageswari, R. U. (2024). Golden Search Optimization based adaptive and diagonal kernel convolution neural network for disease prediction and securing IoT data in cloud. *Applied Soft Computing*, *151*, 111137.

[28] Lin, CJ, Lin, XQ ve Jhang, JY (2024). Akciğer Kanseri Görüntüleri için Karşıt Öğrenme Tabanlı Taguchi Evrişimli Bulanık Sinir Sınıflandırıcı. *IEEE Erişimi* .

[29] Arun Shalin, L. V., Kalaivani, K., Vishal Ratansing, P., & Sathish Kumar, J. (2022). Prognostic Analysis and Early Prediction of Lung Cancer through Gene Analysis. Cybernetics and Systems, 1-17.

[30] Jassim, M. M., & Jaber, M. M. (2022). Hybrid Selection Framework for Class Balancing Approaches Based on Integrated CNC and Decision Making Techniques for Lung Cancer Diagnosis. Eastern-European Journal of Enterprise Technologies, 118(9).

[31] Sinthia, P., Malathi, M., K, A., & Suresh Anand, M. (2022). Improving lung cancer detection using faster region-based convolutional neural network aided with fuzzy butterfly optimization algorithm. Concurrency and Computation: Practice and Experience, 34(27), e7251.

[32] Wu, X., Denise, B. B., Zhan, F. B., & Zhang, J. (2022). Determining association between lung cancer mortality worldwide and risk factors using fuzzy inference modeling and random forest modeling. International Journal of Environmental Research and Public Health, 19(21), 14161.

[33] Geetha, N. (2022). Deep Separable Convolution Network for Prediction of Lung Diseases from X-rays. International Journal of Advanced Computer Science and Applications, 13(6).

[34] Dev, P. P., Patil, S. D., Hulipalled, V. R., & Patil, K. (2022). Fuzzy Sematic Segmentation and Efficient Classification of Lung Cancer Multi-Dimensional Datasets. International Journal of Fuzzy System Applications (IJFSA), 11(3), 1-12.

[35] Hussain, L., Aziz, W., Alshdadi, A. A., Nadeem, M. S. A., & Khan, I. R. (2019). Analyzing the dynamics of lung cancer imaging data using refined fuzzy entropy methods by extracting different features. IEEE Access, 7, 64704-64721.

[36] Arunkumar, C., & Ramakrishnan, S. (2019). Prediction of cancer using customised fuzzy rough machine learning approaches. Healthcare technology letters, 6(1), 13-18.

[37] Polat, K., & Güneş, S. (2008). Principles component analysis, fuzzy weighting pre-processing and artificial immune recognition system based diagnostic system for diagnosis of lung cancer. Expert Systems with Applications, 34(1), 214-221.

[38] Xing, J., Li, C., Wu, P., Cai, X., & Ouyang, J. (2024). Optimized fuzzy K-nearest neighbor approach for accurate lung cancer prediction based on radial endobronchial ultrasonography. Computers in Biology and Medicine, 171, 108038.

[39] Zhang, M., Zhu, L., Sun, Y., Niu, D., & Liu, J. (2022). Computed tomography of ground glass nodule image based on fuzzy C-means clustering algorithm to predict invasion of pulmonary adenocarcinoma. Journal of Radiation Research and Applied Sciences, 15(1), 152-158

[40] Nivedita, A. S., & Sharma, M. K. (2021). Fuzzy mathematical inference system and its application in the diagnosis of lung cancer. Int. J. Agric. Stat. Sci, 17(2), 709-717.

[41] Zakaria, A. S., Shafi, M. A., Mohd Zim, M. A., & Musa, A. N. (2024). Comparison of prediction fuzzy modeling towards high-risk symptoms of lung cancer. Journal of Intelligent & Fuzzy Systems, (Preprint), 1-10.

[42] Manikandan, T., & Bharathi, N. (2017). Hybrid neuro-fuzzy system for prediction of stages of lung cancer based on the observed symptom values. Biomed Res, 28(2), 588-593.

[43] Navaneethakrishnan, M., Anand, M. V., Vasavi, G., & Rani, V. V. (2023). Deep Fuzzy SegNet-based lung nodule segmentation and optimized deep learning for lung cancer detection. Pattern Analysis and Applications, 26(3), 1143-1159.

[44] Khalil, A. M., Li, S. G., Lin, Y., Li, H. X., & Ma, S. G. (2020). A new expert system in prediction of lung cancer disease based on fuzzy soft sets. Soft Computing, 24(18), 14179-14207.

[45] Singh, Y., & Susan, S. (2023). Lung cancer subtyping from gene expression data using general and enhanced fuzzy min–max neural networks. Engineering Reports, 5(11), e12663.

[46] Ramkumar, M., Kalirajan, K., Kumar, U. P., & Surya, P. (2024). Deep volcanic residual U-Net for nodal metastasis (Nmet) identification from lung cancer. Biomedical Engineering Letters, 14(2), 221-233.

[47] Zakaria, A. S., Shafi, M. A., Zim, M. A. M., & Razali, S. N. A. M. (2023). The Use of Fuzzy Linear Regression Modeling to Predict High-risk Symptoms of Lung Cancer in Malaysia. International Journal of Advanced Computer Science and Applications, 14(5).

[48] Nan, Y., Del Ser, J., Tang, Z., Tang, P., Xing, X., Fang, Y., ... & Yang, G. (2023). Fuzzy attention neural network to tackle discontinuity in airway segmentation. IEEE Transactions on Neural Networks and Learning Systems.

[49] Gugulothu, V. K., & Balaji, S. (2023). An automatic classification of pulmonary nodules for lung cancer diagnosis using novel LLXcepNN classifier. Journal of Cancer Research and Clinical Oncology, 149(9), 6049-6057.

[50] Nagaraja, P., & Chennupati, S. K. (2023). Integration of adaptive segmentation with heuristic-aided novel ensemble-based deep learning model for lung cancer detection using CT images. Intelligent Decision Technologies, 17(4), 1135-1160.

[51] Albert Jerome, S., Vijila Rani, K., Mithra, K. S., & Eugine Prince, M. (2023). Watershed segmentation with CAFIS and RCNN classification for pulmonary nodule detection. IETE Journal of Research, 69(8), 5052-5063.

[52] Geng, M., Geng, M., Wei, R., & Chen, M. (2022). Artificial intelligence neural network analysis and application of CT imaging features to predict lymph node metastasis in non-small cell lung cancer. Journal of Thoracic Disease, 14(11), 4384.

[53] Prasad, J., Chakravarty, S., & Krishna, M. V. (2022). Lung cancer detection using an integration of fuzzy K-means clustering and deep learning techniques for CT lung images. Bulletin of the Polish Academy of Sciences Technical Sciences, e139006-e139006.

[54] Thamiselvan, P. (2022). Lung Cancer Examination and Risk Severity Prediction using Data Mining Algorithms. International Journal of Intelligent Systems and Applications in Engineering, 10(3), 737-742.

[55] Nivedita, A. S., & Sharma, M. K. (2021). Fuzzy mathematical inference system and its application in the diagnosis of lung cancer. Int. J. Agric. Stat. Sci, 17(2), 709-717.

[56] Lavanya, M., Kannan, P. M., & Arivalagan, M. (2021). Lung cancer diagnosis and staging using firefly algorithm fuzzy C-means segmentation and support vector machine classification of lung nodules. International Journal of Biomedical Engineering and Technology, 37(2), 185-200.

[57] Priyadharshini, A., & Chitra, S. (2022). A new systematic model for analysis and a hybrid fuzzy multimodality model for lung tumor prediction. Journal of Intelligent & Fuzzy Systems, 42(6), 5591-5604.

[58] Deepa, P., & Suganthi, M. (2020). A fuzzy shape representation of a segmented vessel tree and kernel-induced random forest classifier for the efficient prediction of lung cancer. The Journal of Supercomputing, 76(8), 5801-5824.

[59] Zhao, Z., Zhao, J., Song, K., Hussain, A., Du, Q., Dong, Y., ... & Yang, X. (2020). Joint DBN and Fuzzy C-Means unsupervised deep clustering for lung cancer patient stratification. Engineering Applications of Artificial Intelligence, 91, 103571.

[60] Yazdani, H., Cheng, L. L., Christiani, D. C., & Yazdani, A. (2018). Bounded fuzzy possibilistic method reveals information about lung cancer through analysis of metabolomics. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 17(2), 526-535.

[61] Liao, H., Long, Y., Tang, M., Streimikiene, D., & Lev, B. (2019). Early lung cancer screening using double normalization-based multi-aggregation (DNMA) and Delphi methods with hesitant fuzzy information. Computers & Industrial Engineering, 136, 453-463.

[62] Moitra, D., & Mandal, R. K. (2019). Automated grading of non-small cell lung cancer by fuzzy rough nearest neighbour method. Network Modeling Analysis in Health Informatics and Bioinformatics, 8, 1-9.

[63] Reddy, U., Reddy, B., & Reddy, B. (2019). Recognition of Lung Cancer Using Machine Learning Mechanisms with Fuzzy Neural Networks. Traitement du Signal, 36(1), 87-91.

[64] Palani, D., & Venkatalakshmi, K. (2019). An IoT based predictive modelling for predicting lung cancer using fuzzy cluster based segmentation and classification. Journal of Medical Systems, 43, 1-12.

[65] Arunkumar, C., & Ramakrishnan, S. (2019). Prediction of cancer using customised fuzzy rough machine learning approaches. Healthcare technology letters, 6(1), 13-18.

[66] Yılmaz, A., Arı, S., & Kocabıçak, Ü. (2016). Risk analysis of lung cancer and effects of stress level on cancer risk through neuro-fuzzy model. Computer methods and programs in biomedicine, 137, 35-46.

[67] Manikandan, T., & Bharathi, N. (2016). Lung cancer detection using fuzzy auto-seed cluster means morphological segmentation and SVM classifier. Journal of medical systems, 40, 1-9.

[68] Sakthivel, K., Jayanthiladevi, A., & Kavitha, C. (2016). Automatic detection of lung cancer nodules by employing intelligent fuzzy c-means and support vector machine. BIOMEDICAL RESEARCH-INDIA, 27, S123-S127.

[69] Ghosh, A., & De, R. K. (2016). Fuzzy Correlated Association Mining: Selecting altered associations among the genes, and some possible marker genes mediating certain cancers. Applied Soft Computing, 38, 587-605.

[70] Phillips, M., Altorki, N., Austin, J. H., Cameron, R. B., Cataneo, R. N., Greenberg, J., ... & Schmitt, P. (2007). Prediction of lung cancer using volatile biomarkers in breath. Cancer biomarkers, 3(2), 95-109.

[71] Turna, A., Mercan, C. A., & Bedirhan, M. A. (2005). Prediction of morbidity after lung resection in patients with lung cancer using fuzzy logic. The Thoracic and Cardiovascular Surgeon, 53(06), 368-374.

[72] Page,M.J.,McKenzie,J.E.,Bossuyt,P.M.,Boutron,I.,Hoffmann,T.C.,Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan,S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A.,Lalu,M.M.,Li,T.,Loder,E.W.,Mayo-Wilson,E.,McDonald,S.,McGuinness, LA, & Moher, D. (2021). The PRISMA 2020 state-ment: An updated guideline for reporting systematic reviews. BMJ,372, n71. https://doi.org/10.1136/bmj.n71.

Acknowledgement

Generative AI was used for the enhancement of English language and translation purposes

Appendix

Appendix-1. Features of included studies

No	Research Title	Fuzzy Logic Techniques	Sample Size	Hybrid or Nonhybrid		Fuzzy Using Aim	Methodological Focus	Limits in Research	Performance Metrics
1	Majumder et al., 2024 [26]	Fuzzy Rule-Based Systems Fuzzy C-Means Clustering Fuzzy Inference Systems (FIS): Adaptive Neuro-Fuzzy Inference Systems (ANFIS):	1142 CT scan	Non-hybrid	Mitscherlich function-based fuzzy ranking approach	Assess the uncertainty	Image Classification	Data limitations: does not fully represent real world cases. model complexity, potential for overfitting: dataset is small. No medical confirmation	Accuracy: Q-OTHNCCD dataset**:** 99.54% LIDC-IDRI dataset**:** 95.75% Precision**:** 99.62% Recall**:** 98.61% F1-Score**:** 99.10%
2	Lin et al., 2024 [28]	Fuzzy Neural Networks Taguchi Methods Adversarial Learning	SPIE-AAPM dataset (22.489) and LIDC-IDRI (16.471)	Hybrid	Fuzzy neural classifier (FNC) + convolutional neural network (CNN)	Classification (handling uncertainty)	Image Classification	Data limitations: dataset samples have little samples Method limitations No medical confirmation	AL-TCFNC Accuracy: 88.69% Specificity: 90.00% F1-Score: 89.02% These results are taken from the reported data based on metrics such as accuracy, sensitivity, specificity, and F1-score of the classification experiments performed by the model on SPIE-AAPM and LIDC-IDRI datasets.
3	Kumar et al., 2024 [27]	Fuzzy Logic Controllers Fuzzy Inference Systems Adaptive Fuzzy Systems	COVID-19 Detection X-Ray and Lung Cancer (284 instances)	Non-hybrid	Fuzzy TOPSIS	Multi-criteria decision-making (MCDM) to handle uncertainty and imprecision associated with decision-making processes.	Data Security & Disease Prediction	Data limitations: dataset samples have little samples Method Limitations: The performance of the proposed DKCNN-AK model was only tested on these datasets. No medical confirmation	Accuracy: 98% Sensitivity: 97% (for lung cancer) Specificity: 97% Response Time: 60 seconds Network Capacity: 100 kbps

No	Research Title	Fuzzy Logic Techniques	Sample Size	Hybrid or Nonhybrid		Fuzzy Using Aim	Methodological Focus	Limits in Research	Performance Metrics
4	Ramkumar et al., 2024 [46]	Fuzzy Clustering Fuzzy Rule-Based Systems Fuzzy Inference Systems	126 patients	Non-Hybrid	Fuzzy Clipped Inference System (FCIS)	It helps in accurately detecting nodal metastasis (Nmet) and non-nodal metastasis (Non-Nmet)	Image Classification	Data limitations: dataset samples have little samples Method Limitations: Although the proposed Deep Volcanic Residual U-Net (DVR U-Net) model provides high accuracy, it has only been tested on specific datasets and its applicability for more diverse clinical scenarios is uncertain No medical confirmation	Accuracy: 99.6% (LIDC/IDRI), 98.6% (dual-energy CT data set) Sensitivity: 99.6% (LIDC/IDRI), 97.9% (dual-energy CT data set) Specificity: 98.6% (LIDC/IDRI), 99.5% (dual-energy CT data set)
5	Zakaria et al., 2024 [41]	Fuzzy Rule-Based Systems Fuzzy Inference Systems (FIS) Fuzzy Clustering Adaptive Neuro-Fuzzy Inference Systems (ANFIS)	124 lung cancer patients	Non-hybrid	Fuzzy linear regression	Managing uncertainty	Risk Prediction	Data limitations: - Method Limitations: - No medical confirmation	
6	Xing et al., 2024 [38]	Fuzzy K-Nearest Neighbor (K-NN): Fuzzy Clustering: Fuzzy Inference Systems (FIS):	156 patients	Hybrid	Fuzzy K-Nearest Neighbor (FKNN) + Enhanced Manta Ray Foraging Optimization (ECMRFO)	Classification (improve the classification accuracy), managing uncertainty	Image Classification	Data limitations: The dataset used in the study is limited to 156 patients. This dataset has not been tested in larger and different clinical scenarios. Method Limitations: The model was only tested on a specific dataset and not tested with other imaging techniques or data sets. No medical confirmation	Accuracy: 99.38 % Sensitivity: 100% Specificity: 98.89% F1-Score: 99.33%

No	Research Title	Fuzzy Logic Techniques	Sample Size	Hybrid or Nonhybrid		Fuzzy Using Aim	Methodological Focus	Limits in Research	Performance Metrics
6	Xing et al., 2024 [38]	Fuzzy K-Nearest Neighbor (K-NN): Fuzzy Clustering: Fuzzy Inference Systems (FIS):	156 patients	Hybrid	Fuzzy K-Nearest Neighbor (FKNN) + Enhanced Manta Ray Foraging Optimization (ECMRFO)	Classification (improve the classification accuracy), managing uncertainty	Image Classification	Data limitations: The dataset used in the study is limited to 156 patients. This dataset has not been tested in larger and different clinical scenarios. Method Limitations: The model was only tested on a specific dataset and not tested with other imaging techniques or data sets. No medical confirmation	Accuracy: 99.38 % Sensitivity: 100% Specificity: 98.89% F1-Score: 99.33%
7	Zakaria et al., 2023 [47]	Fuzzy Linear Regression Fuzzy Rule-Based Systems Fuzzy Inference Systems (FIS)	165 lung cancer patients	Non-hybrid	Handle uncertainty in predicting high-risk symptoms of lung cancer	Fuzzy linear regression.	Risk Prediction	Data limitations: The dataset used in the study is limited to 124 samples. Method Limitations: The model has not been tested on different datasets and clinical scenarios. It is also limited in the detection of early-stage lung cancer Symptoms No medical confirmation and since the dataset consist of advanced lung cancer cases, there are deficiencies in early-stage detection	Mean Square Error (MSE): 1.455 (H=0.0) Root Mean Square Error (RMSE): 1.206 (H=0.0) Highest Risk Symptoms: Hemoptysis (14.5494) and chest pain (10.6765) were found to be the highest risk symptoms

No	Research Title	Fuzzy Logic Techniques	Sample Size	Hybrid or Nonhybrid		Fuzzy Using Aim	Methodological Focus	Limits in Research	Performance Metrics
8	Nan et al., 2023 [48]	Fuzzy Attention Mechanism Fuzzy Rule-Based Systems Fuzzy Clustering Fuzzy Inference Systems (FIS)	130 cases	Hybrid	Reduce uncertainty	Fuzzy logic with Neural network	Image Segmentation	Data limitations: The small size of the data sets may limit the generalizability of the model. Method Limitations: The generalizability of the model across multiple disease states (fibrosis and COVID-19) is limited No medical confirmation	IoU (Intersection over Union): 87.38% (BAS data), 92.22% (COVID-19 data), 82.69% (Fibrosis data) Precision: 91.87% (BAS data), 94.31% (COVID-19 data), 89.04% (Fibrosis data) Detected Branch Ratio (DBR): 89.01% (BAS data), 90.18% (COVID-19 data), 73.44% (Fibrosis data) Detected Length Ratio (DLR): 92.71% (BAS data), 93.30% (COVID-19 data), 78.98% (Fibrosis data)
9	Singh & Susan, 2023 [45]	Fuzzy Min-Max Neural Networks (FMMNN): Enhanced FMMNN: Fuzzy Rule-Based Systems:	203 samples	Hybrid	General Fuzzy min-max (GFMM) and Enhanced Fuzzy min-max (EFMM) neural networks	Classification (classifying lung cancer subtypes) and managing uncertainty	Gene Expression Analysis	Data limitations: The dataset used in the study is limited to 203 samples. Imbalance in the data set can also affect the classification results. Method Limitations: It was observed that EFMM was not effective in small sample sizes. No medical confirmation	Accuracy: 98.04% (Validation), 94.06% (Cross-validation). Cross-validation Accuracy: 94.06 EFMM Accuracy: 90.2% (Validation), 93.07% (Cross-validation) Execution Time: 4.57 seconds (GFMM)

No	Research Title	Fuzzy Logic Techniques	Sample Size	Hybrid or Nonhybrid		Fuzzy Using Aim	Methodological Focus	Limits in Research	Performance Metrics
10	Chang, 2023 [13]	Fuzzy Evaluation Systems: Fuzzy Inference Systems (FIS): Fuzzy Rule-Based Systems:	22,489 lung cancer CT images	Non-hybrid	Fuzzy evaluation approach	Handle uncertainty	Image Classification	Data limitations: The dataset may limit the generalizability of the model to other clinical scenarios as it focuses only on a specific type of cancer Method Limitations: The performance of the model on other imaging techniques and multi-class classification problems has not been evaluated. This limits its success in more complex scenarios. No medical confirmation	Accuracy: 99.19% (optimized LeNet-5 CNN) Sensitivity: 99.80% Specificity: 98.60% RPI (Recognition Performance Index): 1.0496 (indicates superiority of the optimized model)
11	Navaneeth akrishnan et al., 2023 [43]	Fuzzy SegNet Fuzzy Clustering Fuzzy Inference Systems (FIS) Optimized Deep Learning	1018 Lung CT Scan	Hybrid	Fuzzy C-Means clustering and a Deep Fuzzy SegNet	Clustering	Image Segmentation	Data limitations: Limited number of chest CT images Method Limitations: It is unclear how the model will perform on different datasets and other cancer types. No medical confirmation	Accuracy: 92.43 Sensitivity: 94.21 Specificity: 89.15
12	Gugulothu & Balaji, 2023 [49]	LLXcepNN Classifier Fuzzy Rule-Based Systems Fuzzy Clustering Fuzzy Inference Systems (FIS)	1010 patients	Hybrid	Geodesic Fuzzy C-Means Clustering (GFCM) + Deep Learning models	Classification	Image Classification	Data limitations: The dataset is limited to low-resolution lung tomography images. Method Limitations: The LLXcepNN model has not been tested for applicability in different clinical scenarios and more complex cases. No medical confirmation	Accuracy: 96.89 % Sensitivity: 95.98 % Specificity: 96.78 % Error Rate: 3.12 %

No	Research Title	Fuzzy Logic Techniques	Sample Size	Hybrid or Nonhybrid		Fuzzy Using Aim	Methodological Focus	Limits in Research	Performance Metrics
13	Nagaraja & Chennupati , 2023 [50]	Adaptive Segmentation: Fuzzy Rule-Based Systems: Heuristic-Aided Ensemble Learning: Fuzzy Inference Systems (FIS)	2000 CT images	Hybrid	Adaptive fuzzy clustering + Improved Harris Hawks Optimization (IHHO) algorithm	Classification and managing uncertainties	Image Classification	Data limitations: - Method Limitations: - No medical confirmation	-
14	Lin & Yang, 2023 [21]	Convolutional Fuzzy Neural Network (F-CFNN)	22.489 Lung Cancer CT İmages	Hybrid	Fuzzy logic and deep learning	Classification and managing uncertainties	Image Classification	Data limitations: The SPIE-AAPM dataset contains a limited number of images and has not been tested in different populations. Method Limitations: The Taguchi-based F-CFNN model was optimized only on this data set. No medical confirmation	Accuracy: 99.98 % Sensitivity: 100% Specificity: 99.96%
15	Albert Jerome et al., 2023 [39]	Fuzzy bean-based classifier for medical image classification and a classifier optimized using fuzzy texture segmentation rules for image segmentation. Coactive adaptive neuro-fuzzy interference system classifier (CAFIS)	455 patients	Hybrid	Coactive Adaptive Neuro-Fuzzy Inference System (CAFIS) + Recurrent Convolutional Neural Network (RCNN)	Classification	Image Classification	Data limitations: LIDC-IDRI and tested on a limited clinical dataset. Method Limitations: WSBTI segmentation algorithm and RCNN classifiers were tested only on specific datasets No medical confirmation	Accuracy: 97.6% (RCNN) Sensitivity: 97.0% (CAFIS) Specificity: 97.6% (RCNN)

No	Research Title	Fuzzy Logic Techniques	Sample Size	Hybrid or Nonhybrid		Fuzzy Using Aim	Methodological Focus	Limits in Research	Performance Metrics
16	Shalin et al., 2022 [29]	Fuzzy Inference Systems (FIS) Fuzzy Rule-Based Systems Fuzzy Clustering	12.533 gene expression	Non-Hybrid	Feature Variance Based Fuzzy Classifier (FVFC)	Classification	Gene Expression Analysis	Data limitations: Human GeneAtlas dataset contains limited gene Expression No medical confirmation	Accuracy: 92.27% Precision: 0.8971% Recall: 0.8838%
17	Geng et al., 2022 [52]	Fuzzy Inference Systems (FIS) Fuzzy Rule-Based Systems Fuzzy Clustering	82 patients	Non-Hybrid	Fuzzy enhancement algorithm	Enhance image contrast	Image Classification	Data limitations: The dataset used is limited to a limited group of patients and a specific hospital. Method Limitations: The method used in the model is optimized only with certain parameters and does not include different datasets or genetic variables No medical confirmation	Accuracy: 95.1% Sensitivity: 90.9% Specificity: 100%
18	Wu et al., 2022 [32]	Fuzzy Inference System Fuzzy Rule-Based Systems: Fuzzy Clustering:	1097 observations	Non-Hybrid	Fuzzy inference modeling (FIM)	Classification and managing uncertainties	Epidemiological Analysis	Data limitations: The dataset was analyzed based only on specific risk factors and limited geographical areas, which limits its general validity. Method Limitations: The performance of the modeling methods has not been tested on larger datasets that can be generalized. No medical confirmation	Random Forest Model: 96.17% accuracy rate was achieved

No	Research Title	Fuzzy Logic Techniques	Sample Size	Hybrid or Nonhybrid		Fuzzy Using Aim	Methodological Focus	Limits in Research	Performance Metrics
19	Sinthia et al., 2022 [31]	Fuzzy Butterfly Optimization Algorithm: Fuzzy Rule-Based Systems: Fuzzy Inference Systems (FIS):	47 participants	Hybrid	Fuzzy Butterfly Optimization Algorithm (FBOA) + Faster RCNN	Classification in decision making process	Image Classification	Data limitations: Limited sample size and therefore limited generalizability and applicability of the model to larger patient populations Method Limitations: Although the proposed RCNN and fuzzy butterfly optimization algorithm provides high accuracy, the training process and computational costs of the model are high. Moreover, the algorithm has not been tested with different datasets. No medical confirmation	Accuracy: 97% Sensitivity: 98% F1-Score: 99%
20	Prasad et al., 2022 [53]	Fuzzy K-Means Clustering Fuzzy Rule-Based Systems Fuzzy Inference Systems (FIS)	1018 cases	Hybrid	Fuzzy K-means clustering with deep learning techniques	Classification	Image Classification	Data limitations: It was performed with a limited dataset and the performance of the model on different populations and clinical scenarios is uncertain Method Limitations: K-means clustering and deep learning methods are optimized only on a specific dataset No medical confirmation	Accuracy: 96% Sensitivity: 99 Specificity: 100%

No	Research Title	Fuzzy Logic Techniques	Sample Size	Hybrid or Nonhybrid		Fuzzy Using Aim	Methodological Focus	Limits in Research	Performance Metrics
21	Zhang et al., 2022 [39]	Fuzzy C-Means Clustering Fuzzy Rule-Based Systems Fuzzy Inference Systems (FIS)	65 cases	Non-hybrid	Fuzzy C-Means (FCM) clustering algorithm	Classification	Image Classification	Data limitations: It Limited sample (65) may limit the generalizability of the model to large patient populations Method Limitations: The Fuzzy C-Means clustering algorithm has only been tested on a specific dataset and has not been validated with other datasets No medical confirmation	Accuracy: 77.8% Specificity: 75.0%
22	Jassim & Jaber, 2022 [30]	Fuzzy Decision-Making Techniques Fuzzy Rule-Based Systems Fuzzy Inference Systems (FIS)	IQ-OTH/NCCD lung cancer (1097 samples)	Hybrid	Fuzzy Multicriteria Decision Making + Deep Learning Techniques	Classification	Data Imbalance Handling	Data limitations: The study was performed with a limited number of lung cancer images, Method Limitations: The proposed deep learning model was tested with certain limited data sets No medical confirmation	Accuracy: 99.27% Sensitivity: 99.33% Specificity: 99%
23	Geetha & Joseph, 2022 [33]	Fuzzy Inference Systems (FIS) Fuzzy Rule-Based Systems Fuzzy Clustering	5,606 X-ray images (used %5)	Hybrid	Enriched Auto-Seed Fuzzy Means Morphological Clustering (EASFMC) + Modified Butterfly Optimization Algorithm (MBOA)	Classification	Image Classification	Data limitations: Limited to NIH Chest X-ray dataset Method Limitations: The model is only tested on this dataset No medical confirmation	Accuracy: 98.45% Sensitivity: 95% F1-Score: 98.85%

No	Research Title	Fuzzy Logic Techniques	Sample Size	Hybrid or Nonhybrid		Fuzzy Using Aim	Methodological Focus	Limits in Research	Performance Metrics
24	Dev et al., 2022 [34]	Fuzzy Semantic Segmentation Fuzzy Rule-Based Systems Fuzzy Inference Systems (FIS)	150 lung cancer patient records	Hybrid	Fuzzy Semantic Segmentation technique + Convolutional Neural Networks (CNNs)	Classification and managing uncertainties	Image Classification	Data limitations: Data consists of a limited number of images Method Limitations: The proposed fuzzy logic and DNN based methods are tested only on specific data sets, No medical confirmation	Accuracy: 91.42 % Sensitivity: 90.38 % Specificity: 82.41 %
25	Thamilselvan, 2022 [54]	Fuzzy Inference Systems (FIS) Fuzzy Rule-Based Systems Fuzzy Clustering	-	Hybrid	Enhanced K-Nearest Neighbor (EKNN) and Advanced Classification and Regression Tree (ACART)	Managing Uncertainties	Risk Prediction	Data limitations: The study was conducted using a limited dataset, Method Limitations: Although the algorithms used offer high accuracy, they have not been tested with other datasets and the optimization of the algorithms could not be evaluated in different clinical scenarios No medical confirmation	Accuracy: 97% (KNN), 98.3% (ACART)
26	Nivedita et al., 2021 [55]	Fuzzy Mathematical Inference System: Fuzzy Rule-Based Systems: Fuzzy Inference Systems (FIS):	-	Non-hybrid	Mamdani Fuzzy Inference System	Managing uncertainties	Lung Cancer Diagnosis	Data limitations: Limited to data from a specific hospital and a limited number of patients Method Limitations: The fuzzy inference system only works with specific symptoms and does not consider other possible cancer symptoms No medical confirmation	The model was tested using MATLAB and was able to correctly classify different degrees of symptoms, but the exact performance metrics are not specified in the paper.

No	Research Title	Fuzzy Logic Techniques	Sample Size	Hybrid or Nonhybrid		Fuzzy Using Aim	Methodological Focus	Limits in Research	Performance Metrics
27	Lavanya et al., 2021 [56]	Firefly Algorithm Fuzzy C-Means Segmentation Fuzzy Inference Systems (FIS) Fuzzy Rule-Based Systems	-	Hybrid	Firefly Algorithm Fuzzy C-Means (FA-FCM) Segmentation + Support Vector Machine (SVM)	Classification and managing uncertainties	Image Classification	Data limitations: - Method Limitations: - No medical confirmation	
28	Priyadharshini & Zoraida, 2021 [57]	Bat-Inspired Metaheuristic Algorithms Fuzzy Inference Systems (FIS) Fuzzy Rule-Based Systems	1018 cases	Non-Hybrid	Fuzzy C-means (FCM)	Classification	Image Classification	Data limitations: Limited to the LIDC-IDRI dataset Method Limitations: The segmentation and classification processes using the Fuzzy C-Means algorithm and BAT optimization were optimized only on this data set No medical confirmation	Accuracy: 97.43
29	Deepa & Suganthi, 2020 [58]	Fuzzy Shape Representation Kernel-Induced Random Forest Classifier Fuzzy Rule-Based Systems	1018 CT scan images	Hybrid	Fuzzy logic + kernel-induced random forest classifiers	Managing uncertainties	Image Classification	Data limitations: Limited to the LIDC-IDRI dataset. Method Limitations: The fuzzy shape representation and kernel-induced random forest classifier were tested on this dataset only No medical confirmation	Accuracy: 94% Sensitivity (Recall): 91.2% Specificity: 92.4%

No	Research Title	Fuzzy Logic Techniques	Sample Size	Hybrid or Nonhybrid		Fuzzy Using Aim	Methodological Focus	Limits in Research	Performance Metrics
30	Zhao et al., 2020 [59]	Fuzzy C-Means Clustering Deep Belief Networks (DBN) Fuzzy Rule-Based Systems	1018 cases lung CT images	Hybrid	Deep Belief Networks (DBN) + Fuzzy C-Means clustering	Clustering	Clustering	Data limitations: The datasets are from a limited population and have not been tested in broader or different clinical scenarios Method Limitations: The methodology used was tested only on specific data sets. No medical confirmation	Accuracy: 99.19% Sensitivity: 99.80% Specificity: 98.60%
31	Khalil et al., 2020 [44]	Fuzzy Soft Sets Fuzzy Rule-Based Systems Fuzzy Inference Systems (FIS)	45 Patients	Non-hybrid	Fuzzy Soft Expert System	Managing uncertainties	Risk Prediction	Data limitations: The study used a dataset of only 45 test patients, which may not fully represent a diverse population of lung cancer patients. Method Limitations: The proposed fuzzy soft expert system is based on complex fuzzy logic, which may require further validation in larger datasets and real-world clinical settings. No medical confirmation	Accuracy: 100% on test data for lung cancer prediction
32	Yazdani et al., 2020 [60]	Bounded Fuzzy Possibilistic Method Fuzzy Rule-Based Systems Fuzzy Inference Systems (FIS)	231 Samples	Non-hybrid	Bounded Fuzzy Possibilistic Method (BFPM)	Managing uncertainties	Metabolomics Analysis	Data limitations: Limited to the dataset Method Limitations: The Bounded Fuzzy Probability Method (BFPM) requires high computational resources, which may limit its practical application in real-time environments. No medical confirmation	Accuracy: Significant differences were found between serum samples of healthy individuals and serum samples of lung cancer patients, providing insight for early detection and diagnosis.

No	Research Title	uzzy Logic Techniques	Sample Size	Hybrid or Nonhybrid		Fuzzy Using Aim	Methodological Focus	Limits in Research	Performance Metrics
33	Liao et al., 2019 [61]	Hesitant Fuzzy Information Double Normalization-Based Multi-Aggregation (DNMA) Fuzzy Rule-Based Systems	-	Hybrid	DNMA method + the fuzzy Delphi method + hesitant fuzzy information	Managing uncertainties	Early Detection	Data limitations: Limited to the samples Method Limitations: The extended method was optimized only with specific measures and was not tested with other datasets, No medical confirmation	Accuracy: 92.43% Sensitivity: 94.21% Specificity: 89.15%
34	Moitra and Mandal, 2019 [62]	Fuzzy rough nearest neighbour method. Fuzzy Rule-Based Systems. Fuzzy Inference Systems (FIS)	211 Patients	Hybrid	Fuzzy logic + rough set theory and the nearest neighbor approach	Classification and managing uncertainties	Classification	Data limitations: Limited to the samples Method Limitations: Fuzzy Rough Nearest Neighbor cleaning has high programming cost due to its complexity compared to other temperatures. No medical confirmation	Accuracy : 95% Sensitivity (Recall) : 93% F-measure: 93%
35	Reddy and Reddy, 2019 [63]	Frequency Ratio Fuzzy C-Means (FRFCM) Fuzzy C-Means (FCM) Kernelized Fuzzy C-Means (KFCM) Spatially Adaptive Fuzzy C-Means (SAFCM) Fuzzy Local Information C-Means (FLICM)	-	Hybrid	Fuzzy logic + Neural Networks	Classification and Managing uncertainties	Image Classification	Data limitations: Limited number and types of CT images Method Limitations: Neural networks and fuzzy logic methods used in the study were tested only on specific data sets No medical confirmation	Accuracy: 96.67%

No	Research Title	Fuzzy Logic Techniques	Sample Size	Hybrid or Nonhybrid		Fuzzy Using Aim	Methodological Focus	Limits in Research	Performance Metrics
36	Palani & Venkatalakshmi, 2019 [64]	Fuzzy C-Means Clustering algorithm	113 lung images	Hybrid	Fuzzy C-Means Clustering + temporal features + Association Rule Mining (ARM) + Decision Tree (DT) classifiers, and Convolutional Neural Networks (CNN)	Classification and prediction	Prediction Model	Data limitations: Data sets are limited and have not been tested on larger, diverse populations Method Limitations: The proposed IoT-based forecasting model has been evaluated with limited tests and specific datasets. It has not been tested in more complex scenarios. No medical confirmation	Accuracy: 99.54% Sensitivity: 85% Specificity: 85%
37	Hussain et al., 2019 [35]	Refined Fuzzy Entropy Methods Fuzzy Rule-Based Systems Fuzzy Inference Systems (FIS)	76 patients (945 images)	Non-hybrid	Refined Fuzzy Entropy Multiscale Fuzzy Entropy and Refined Composite Multiscale Fuzzy Entropy – just fuzzy techniques	Managing uncertainties	Image Feature Extraction	Data limitations: small sample size Method Limitations: The complexity of extracting features such as texture, morphological and elliptical Fourier descriptors limits the generalizability of the results across various datasets. No medical confirmation	Using different entropy measurements (P values as low as 1.95E-50 for tissue features), high statistical significance was achieved in differentiating NSCLC from SCLC.

No	Research Title	Fuzzy Logic Techniques	Sample Size	Hybrid or Nonhybrid		Fuzzy Using Aim	Methodological Focus	Limits in Research	Performance Metrics
38	Arunkumar et al., 2019 [65]	Fuzzy Rough Sets Fuzzy Rule-Based Systems Fuzzy Inference Systems (FIS)	**Leukemia :**72 **Central Nervous System:**60 **Lung cancer:**181 **Ovarian cancer:**253	Hybrid	Fuzzy rough set theory + customized fuzzy triangular norm operator for feature selection	Improving classification accuracy, enhancing the prediction of cancer types	Cancer Prediction	Data limitations: The study is based on microarray gene expression data and the datasets contain a limited number of samples. This may limit the generalizability of the model across different data sets and real-world applications Method Limitations: The proposed fuzzy rough clustering algorithm was tested only on specific data sets. No medical confirmation	Accuracy (CA): 98.11% (for lung cancer dataset) Precision: 98.1 F1-Score: 98.7
39	Manikandan & Bharathi, 2017 [42]	Hybrid Neuro-Fuzzy System Fuzzy Rule-Based Systems Fuzzy Inference Systems (FIS)	217 samples (167 lung cancer patients, 50 normal patients)	Hybrid	Fuzzy logic and Neural Networks	Managing uncertainty	Lung Cancer Staging	Data limitations: Limited sample size therefore the results cannot be generalized to a wider population Method Limitations: The proposed neural network and fuzzy logic system are tested on a specific dataset, which limits the performance of the model on other datasets. No medical confirmation	Accuracy: 97.7% Sensitivity: 100% Specificity: 80%
40	Yilmaz et al., 2016 [66]	Fuzzy Risk Assessment Models Fuzzy Inference Systems (FIS) Fuzzy Rule-Based Systems	1536 samples	Hybrid	Fuzzy logic and Neural Networks	Managing uncertainty	Risk Analysis	Data limitations: The study is limited to lung cancer data from 1536 people. Method Limitations: The proposed model is optimized only with a specific data set No medical confirmation	Accuracy: 94.64 Sensitivity: 96.69% (for stress model)

No	Research Title	Fuzzy Logic Techniques	Sample Size	Hybrid or Nonhybrid		Fuzzy Using Aim	Methodological Focus	Limits in Research	Performance Metrics
41	Manikandan & Bharathi, 2016 [67]	Fuzzy Auto-Seed Cluster Means SVM Classifier, Fuzzy Rule-Based Systems	106 samples	Hybrid	Fuzzy clustering with a Support Vector Machine (SVM)	Clustering	Image Classification	Data limitations: Limited patient data Method Limitations: The fuzzy automatic seed clusters morphological segmentation algorithm used was tested only with specific datasets and was not validated on different datasets No medical confirmation	Accuracy: 94% Sensitivity: 100% Specificity: 93%
42	Sakthivel et al., 2016 [68]	Intelligent Fuzzy C-Means Clustering SVM Classifier Fuzzy Rule-Based Systems	400 lung CT images	Hybrid	Intelligent Fuzzy C-Means (IFCM) + Support Vector Machine (SVM)	Managing uncertainty	Image Classification	Data limitations: Limited number of CT images Method Limitations: The proposed fuzzy C-means algorithm was tested only on this dataset. No medical confirmation	Accuracy: 97.6% Sensitivity: 98% Specificity: 97%
43	Ghosh & De, 2016 [69]	Fuzzy Correlated Association Mining Fuzzy Rule-Based Systems Fuzzy Inference Systems (FIS)	**Lung cancer;** 86 **Colon cancer;**18 **Breast cancer;**4 tumor **Sarcoma;** 39 **Leukemia;** 43 tumor	Non-hybrid	Fuzzy Correlated Association Mining (FCAM)	Managing uncertainties and Association mining	Gene Association Mining	Data limitations: - Method Limitations: - No medical confirmation	-

No	Research Title	Fuzzy Logic Techniques	Sample Size	Hybrid or Nonhybrid		Fuzzy Using Aim	Methodological Focus	Limits in Research	Performance Metrics
44	Daliri, 2012 [20]	Fuzzy Extreme Learning Machines (FELMs) Genetic Algorithm Fuzzy Rule-Based Systems	32 Samples	Hybrid	Genetic Algorithm for feature selection and a Fuzzy Extreme Learning Machine	Classification	Cancer Diagnosis	Data limitations: Limited patient data Method Limitations: The proposed genetic algorithm and fuzzy logic based learning machine are optimized on specific datasets and not tested on other datasets No medical confirmation	Accuracy: 98.85%
45	Polat and Günes, 2008 [37]	Fuzzy Weighting Pre-Processing Fuzzy Rule-Based Systems Fuzzy Inference Systems (FIS)	32 Samples	Non-hybrid	Fuzzy membership functions	Improve the performance of the Artificial Immune Recognition System (AIRS) classifier	Cancer Diagnosis	Data limitations: Limited dataset Method Limitations: The proposed algorithm has only been tested on a specific dataset No medical confirmation	Accuracy: 100% (with PCA, Fuzzy Weighing, AIRS)
46	Phillips et al., 2007 [70]	Fuzzy Inference Systems (FIS) Fuzzy Rule-Based Systems Fuzzy Clustering	193 subjects 211 controls	Non-hybrid	Fuzz logic model	Enhance the prediction accuracy of lung cancer	Biomarker Analysis	Data limitations: - Method Limitations: - No medical confirmation	-
“	Turna et al., 2005 [71]	Fuzzy Inference Systems (FIS) Fuzzy Rule-Based Systems Fuzzy Clustering	91 patients	Non-hybrid	Fuzzy logic model	Predict the risk of complications after lung resection surgery	Risk Prediction	Data limitations: Limited dataset Method Limitations: The proposed fuzzy logic system was built using a limited number of parameters. Its validity for different clinical scenarios has not been tested. No medical confirmation	Sensitivity: 76% Accuracy: 83%

No	Research Title	Fuzzy Logic Techniques	Sample Size	Hybrid or Nonhybrid		Fuzzy Using Aim	Methodological Focus	Limits in Research	Performance Metrics
48	Schneider et al., 2003 [10]	Fuzzy Logic-Based Tumor Marker Profiles Fuzzy Rule-Based Systems Fuzzy Inference Systems (FIS)	33 lung cancer patient	Non-hybrid	Fuzzy logic rule-based system	Improve the sensitivity of detecting tumor	Cancer Progression Analysis	Data limitations: - Method Limitations: - No medical confirmation	-
49	Schneider et al., 2002 [9]	Fuzzy Logic-Based Tumor-Marker Profiles Fuzzy Rule-Based Systems Fuzzy Inference Systems (FIS)	175 lung cancer patients	Non-hybrid	Fuzzy logic system	Enhance the diagnostic sensitivity of tumor markers for lung cancer	Cancer Detection	Data limitations: Limited dataset Method Limitations: The proposed fuzzy logic-based model was tested only in specific conditions and was not optimized for different patient groups or disease stages No medical confirmation	Accuracy: 92% Sensitivity: 92%