Iranian Journal of Fuzzy Systems

Volume 22, Number 5, (2025), pp. 159-179

Original paper

University of Sistan and Baluchestan

Pneumonia detection in chest X-ray images using Convolutional Neural Network and fuzzy VIKOR

L. Yousofvand [1], M. B. Dowlatshahi [2] and M. Pirdadeh Beiranvand [3]

[1,2,3] *Department of Computer Engineering, Lorestan University, Khorramabad, 68135-1911, Lorestan, Iran*

yousofvand.l@lu.ac.ir,dowlatshahi.mb@lu.ac.ir,pirdadeh@yahoo.com

Abstract

Pneumonia is a life-threatening respiratory disease that requires early and accurate diagnosis for effective treatment and reduced complications. However, conventional diagnostic methods such as PCR are often time-consuming, equipment-dependent, and limited to specialized medical centers. This study introduces a novel anomaly detection framework for pneumonia diagnosis using advanced machine learning techniques applied to chest X-ray images. To enhance classification performance, the framework integrates several feature selection methods, including Correlation-based Feature Selection (CFS) to evaluate feature relevance, Fisher Score to rank features based on discriminative power, Maximum Information Coefficient (MIC) to capture complex dependencies, and Local Learning-based Correlation Feature Selection (LLCFS) to improve accuracy by considering local feature correlations. To further enhance classification performance, this study introduces the first-ever application of Fuzzy VIKOR in pneumonia detection. This fuzzy logic-based ensemble method effectively handles uncertainty in medical imaging data, leading to more balanced decision-making when dealing with conflicting information. The proposed model was trained on a chest X-ray dataset and evaluated using key classification metrics, including accuracy, recall, precision, and F1-score. Experimental results confirm that the model outperforms baseline methods across all metrics, achieving an accuracy of 98.34%. These findings validate the effectiveness of the proposed framework and highlight its high potential for real-world deployment in AI-driven computer-aided diagnosis (CAD) systems, particularly in hospitals and telemedicine applications.

Keywords: Pneumonia detection, anomaly detection, fuzzy MCDM, ensemble feature selection, machine learning.

1 Introduction

Respiratory diseases are among the most significant global health challenges, affecting millions of people annually and contributing substantially to morbidity and mortality rates. Among these diseases, pneumonia is one of the most prominent pulmonary diseases worldwide, which, if not diagnosed and treated in a timely manner, can lead to irreparable complications [19]. Therefore, rapid and accurate diagnosis of these diseases is of paramount importance. In this regard, medical imaging and automated diagnostic methods have increasingly attracted attention. Despite significant advancements in these fields, there are still many challenges in diagnosing diseases from medical images that affect the accuracy and efficiency of these systems [28].

One of the major challenges in this area is data imbalance. In many medical datasets, especially in pneumonia detection, the ratio of normal samples to sick samples is not equal. This issue can lead to biases in predictions and decreased model sensitivity in identifying rare cases [36]. Additionally, high-dimensional data and the complexities present in medical images are other challenges that can lead to overfitting of models. This results in models performing poorly when faced with new, unseen data and being unable to identify new patterns [24]. Furthermore, feature selection is another

Corresponding Author: M. B. Dowlatshahi
Received: April 2025; Revised: July 2025; Accepted: September 2025.
https://doi.org/10.22111/ijfs.2025.51603.9119

significant challenge in processing medical images. Many traditional feature selection methods are not effective in accurately identifying key features and eliminating irrelevant ones, which leads to decreased model accuracy and efficiency.

Despite these challenges, in recent years, machine learning (ML) and deep learning (DL) techniques have shown significant advancements in analyzing medical images and detecting diseases. Models like Convolutional Neural Networks (CNN), Support Vector Machines (SVM), and Decision Trees (DT) have performed excellently in identifying abnormalities in medical images, providing high diagnostic accuracy. However, there are still limitations in these models that need improvement. One of these limitations is the data scale issue and the complexities present in medical images, which require optimization in feature selection and the use of newer methods to reduce these complexities.

Due to the ambiguity and complexity of chest X-rays, relying on a single technique may not ensure both accuracy and reliable decisions. While CNNs effectively extract deep features, they can struggle with uncertainty and data imbalance. To address this, we integrate CNNs with the fuzzy VIKOR method-a multi-criteria decision-making approach suited for handling imprecise and conflicting information. This hybrid framework combines strong feature learning with interpretable, balanced decision-making, making it well-suited for clinical diagnostics.

The aim of this paper is to address these issues and present an innovative approach that, by combining advanced feature selection techniques and machine learning algorithms, improves the detection of respiratory diseases. In this research, various feature selection techniques such as CFS [27], Fisher Score [7], LLCFS [40], and MIC [6] have been employed to select and optimize features and reduce the dimensionality of the data. This allows the model to achieve higher accuracy in disease detection and avoid bias caused by data imbalance. Additionally, the SVM algorithm is used for classification, which performs well in imbalanced datasets and provides a more accurate evaluation of diagnostic models.

In this research, Fuzzy VIKOR, a multi-criteria decision-making technique, is used to optimally combine the selected features, ensuring a good balance between dimensionality reduction and classification accuracy. This method also uses LLCFS and MIC to eliminate redundant and weakly correlated features, enhancing the model's performance when tested on new data and improving detection accuracy.

This research specifically addresses the challenges in detecting respiratory diseases and improves existing methods through an innovative combined approach.

The main contributions of this study are: (1) proposing a novel pneumonia detection framework using chest X-rays with multiple advanced feature selection methods (CFS, Fisher Score, MIC, LLCFS); (2) introducing the first use of Fuzzy VIKOR in medical imaging for balanced decision-making under uncertainty; (3) developing a hybrid model that combines deep learning with traditional classifiers for both accuracy and interpretability; and (4) demonstrating superior performance in experiments, with accuracy up to 98.34% and strong potential for clinical and telemedicine use.

In the following sections of the paper, an overview of previous research and current methods in respiratory disease detection will be presented. Then, in Section 3, key concepts related to feature selection and machine learning models will be discussed in detail. In Section 4, the proposed method will be fully explained, and in Section 5, experimental results and model performance evaluation will be provided. Finally, Section 6 will conclude the study and propose directions for future research.

2 Related works

Chest radiography, particularly chest X-rays (CXR), plays a crucial role in detecting pneumonia. However, a major challenge in clinical practice is the shortage of expert radiologists who can efficiently interpret medical images and assess the severity of the disease. This gap highlights the need for automated decision-support systems that can assist radiologists in diagnosing pneumonia more quickly and accurately. Artificial Intelligence (AI), especially deep learning (DL), has shown significant potential in processing large volumes of medical data, offering a promising approach for pneumonia detection.

AI-driven diagnostic models have demonstrated exceptional capability in reducing detection time and enhancing accuracy, particularly for infectious lung diseases such as COVID-19 and bacterial pneumonia [28]. The application of deep learning to medical imaging has been widely explored, with researchers focusing on identifying robust mod-

els for classifying pneumonia from CXR and CT scans. Among various deep learning architectures, Convolutional Neural Networks (CNNs) are the most commonly used techniques for pneumonia detection, achieving high accuracy rates in binary classification tasks (i.e., distinguishing between pneumonia-positive and normal lung conditions) [22, 26].

Recent studies have investigated various machine learning (ML) and deep learning (DL) techniques for pneumonia detection, especially in the context of COVID-19 diagnosis. AI-based models have been reported to achieve impressive accuracy levels. A systematic review of 22 research papers on AI applications in medical imaging found that AI models achieved accuracy rates of 93.7% for CXR and 89.1% for CT scans, demonstrating the effectiveness of AI in pneumonia detection. These studies also emphasized the role of AI in outbreak prediction and disease management, showcasing its broad impact beyond diagnosis[3, 8, 23]. Further advancements in medical AI have led to the development of models like IMITATE, which employs vision-language pretraining (VLP) to improve the alignment of medical images with textual descriptions in clinical reports [2, 23]. This approach bridges the gap between visual and textual data, enhancing the interpretability of AI-based pneumonia detection models. Additionally, pretrained CNN architectures such as ResNet50, ResNet101, and VGG16 have been widely used to classify pneumonia cases with remarkable success. For instance, ResNet50 achieved a detection accuracy of 92.6%, while models like DarkNet and DenseNet have reported classification accuracies exceeding 98% [17, 29, 31, 34].

Another widely used strategy is transfer learning, where deep learning models are pretrained on large-scale datasets (e.g., ImageNet) and then fine-tuned on medical datasets. This technique has proven effective in pneumonia detection, improving precision, recall, and generalization across different datasets [20] .

Beyond model development, research has also focused on fuzzy logic-based preprocessing techniques for medical images, aimed at improving AI model performance. Methods such as fuzzy edge detection and fuzzy histogram equalization have been shown to enhance image quality, making it easier for deep learning models to extract critical features from chest X-rays. Fuzzy edge detection sharpens structural boundaries in images, aiding CNNs in distinguishing pneumonia-affected lung regions, while fuzzy histogram equalization enhances contrast, improving the visibility of abnormalities [16, 25, 30, 37, 39].

Despite the success of CNN-based pneumonia detection models, certain challenges remain. Overfitting is a significant issue, particularly when models are trained on small or specialized medical datasets, leading to poor generalization in real-world scenarios. However, advanced preprocessing techniques such as fuzzy logic help mitigate these limitations by refining image clarity and reducing noise, ensuring models focus on the most informative features. The integration of fuzzy logic with deep learning has thus been shown to improve classification accuracy and reduce the risk of misdiagnosis [1, 28].

Recent advances in other domains offer valuable strategies applicable to medical diagnostics. For example, the Adaptive Probabilistic Network (APN) effectively captured temporal dependencies in arrhythmia classification [4], while multimodal biometric systems improved recognition by fusing heterogeneous data like signatures and fingerprints [9]. Neural network-based methods in edge computing have also enabled efficient real-time classification [32]. Although developed in different fields, these approaches share principles such as decision-level fusion, robust feature modeling, and scalable architectures-concepts that inspire our use of fuzzy decision fusion (Fuzzy VIKOR) and a hybrid learning pipeline to improve accuracy and interpretability in pneumonia detection from chest X-rays.

In summary, AI-based techniques, particularly deep learning have demonstrated immense potential in automating pneumonia diagnosis using chest X-rays and CT scans. Continuous advancements in feature selection, deep learning architectures, transfer learning, and fuzzy logic preprocessing are shaping the future of AI-driven medical diagnostics. While previous studies using CNNs or pretrained models like ResNet and DenseNet have achieved high accuracy, few have explored combining multiple feature selection methods to enhance model robustness. Moreover, the use of fuzzy multi-criteria decision-making methods—particularly Fuzzy VIKOR—has been largely overlooked in medical image analysis.Recent MCDM-based approaches such as MFS-MCDM [11], VMFS [13], Ant Colony MCDM [14], Unsupervised Fuzzy MCDM [5], and AE-MCDM [10] demonstrate that multi-criteria decision-making can effectively guide feature selection, balancing the trade-off between relevance and redundancy. Unlike single-pipeline approaches, our method introduces an ensemble strategy that integrates deep feature extraction with multi-stage feature selection (CFS, Fisher Score, LLCFS, MIC) and Fuzzy VIKOR-based decision fusion, aiming to improve interpretability and diagnostic performance under uncertainty and data imbalance. This study builds upon these previous works by integrating multiple feature selection techniques with fuzzy logic-based ensemble learning, aiming to further enhance pneumonia detection

accuracy.

3 Preliminaries

This section provides an overview of the fundamental concepts related to feature selection and deep learning models used in pneumonia detection. First, we examine the ResNet architecture, followed by an introduction to various feature selection methods that play a key role in enhancing the performance of machine learning models by eliminating irrelevant features and improving classification accuracy. Finally, the VIKOR method is introduced as a multi-criteria decision-making approach for optimal feature selection.

3.1 ResNet (Residual Neural Network)

ResNet (Residual Neural Network) [15] is a powerful deep learning architecture designed to address the vanishing gradient problem in deep networks. Traditional deep models, such as VGGNet [38], suffer from performance degradation as their depth increases, making training inefficient. ResNet overcomes this issue by introducing skip connections (residual connections) that allow information to bypass certain layers, ensuring better gradient flow and faster convergence during training.

The key idea behind ResNet is the use of residual learning, where the network learns the difference (residual) between the input and output of a layer, rather than directly learning a new transformation. This approach has significantly improved performance in image classification, object detection, and medical imaging tasks.

Different versions of ResNet, such as ResNet-18, ResNet-34, ResNet-50, and ResNet-101, vary primarily in the number of layers. In this study, ResNet-50 is employed for feature extraction from chest X-ray images, leveraging its deep hierarchical structure to capture complex patterns in medical data. Figure 1 illustrates the architecture of the ResNet network.

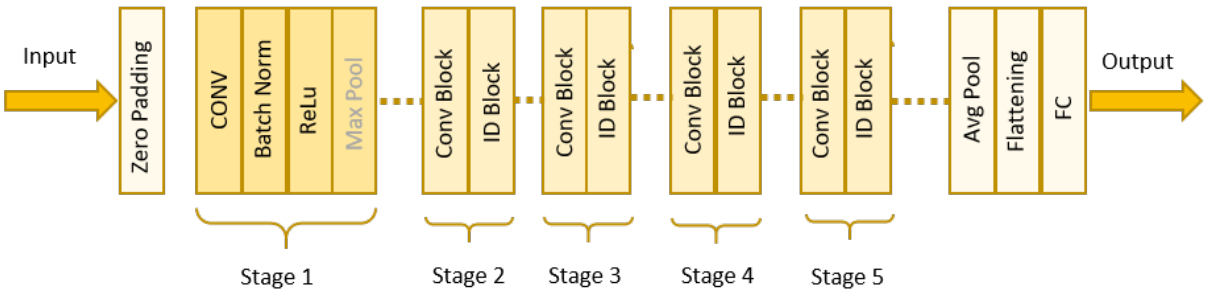


Figure 1: Architecture of the ResNet.

3.2 Feature selection

Feature selection is a critical preprocessing technique in machine learning that helps improve model efficiency by eliminating redundant, irrelevant, or noisy features. This process reduces computational complexity, prevents overfitting, and enhances classification performance. Feature selection methods are generally categorized into three main types:

- **Filter Methods:** Independent of any specific learning algorithm, these methods evaluate features based on statistical metrics such as correlation, mutual information, or variance.
- **Wrapper Methods:** These methods iteratively select the best subset of features by training a learning model and assessing performance improvements.
- **Embedded Methods:** Combining the advantages of filter and wrapper approaches, these methods integrate feature selection within the model training process, dynamically selecting the most relevant features.

3.3 VIKOR (VIseKriterijumska Optimizacija I Kompromisno Resenje)

VIKOR [33] is a multi-criteria decision-making (MCDM) method designed for ranking and selecting the best decision when multiple conflicting criteria are involved. In feature selection, different methods may rank features differently, making it challenging to choose the most relevant ones. VIKOR optimally integrates multiple ranking techniques, ensuring a balanced selection of features for classification.

Steps in the VIKOR Process:

Step 1. Construct the Decision Matrix:
The first step is to create a decision matrix, where each row represents a feature, and each column corresponds to a criterion.

$$D = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}. \tag{1}$$

Where x_{ij} represents the value of criterion j for feature i, and there are m features and n criteria.

Step 2. Normalize the Decision Matrix:
Normalization is used to ensure fair comparisons by converting criteria scores to a uniform scale. This process is especially important when the criteria are different and comparing them is challenging.
To make the values comparable, normalize the matrix using linear normalization:

For beneficial criteria (where higher values are better):

$$f_j^* = \max_i x_{ij}, \quad f_j^- = \min_i x_{ij}, \tag{2}$$

$$f_{ij} = \frac{x_{ij} - f_j^-}{f_j^* - f_j^-}. \tag{3}$$

For cost criteria (where lower values are better):

$$f_j^* = \min_i x_{ij}, \quad f_j^- = \max_i x_{ij}, \tag{4}$$

$$f_{ij} = \frac{f_j^- - x_{ij}}{f_j^- - f_j^*}. \tag{5}$$

Step 3. Determine the Ideal and Worst Values for Each Criterion:
After normalization, the ideal (best) and worst (anti-ideal) values are identified for each criterion. This step is crucial to understand which features are closer to the ideal choice and which ones are farther from it.
Two measures for each criterion are calculated:

$$S_i = \sum_{j=1}^{n} w_j (f_j^* - f_{ij}), \tag{6}$$

where w_j is the weight of criterion j, indicating its importance.

$$R_i = \max_j \left[w_j (f_j^* - f_{ij}) \right]. \tag{7}$$

Step 4. Calculate the Q_i Score for Final Ranking:
The final score Q_i is calculated, which integrates S_i and R_i into a single ranking score using a trade-off parameter (v), typically set to 0.5 to balance between best overall performance and worst-case performance. This score leads to a final ranking of features, where those with the lowest Q_i scores are prioritized for selection.

$$Q_i = v \frac{S_i - S^*}{S^- - S^*} + (1 - v) \frac{R_i - R^*}{R^- - R^*}, \tag{8}$$

where:

$$S^* = \min_i S_i, \quad S^- = \max_i S_i, \quad R^* = \min_i R_i, \quad R^- = \max_i R_i.$$

Step 5. Select the Best Features:
Features with the lowest Q_i values are selected, ensuring optimal classification accuracy. In cases where the best option is not clear, two conditions should be checked:

Acceptance condition:

$$Q(A_2) - Q(A_1) \geq \frac{1}{m-1}, \tag{9}$$

where A_1 and A_2 are two options ranked in the top positions.

Decision condition:
If the first condition is not satisfied, a set of alternatives may be selected for further decision making.

4 Dataset and methodology

Figure 2 illustrates the framework of the proposed method. In this research, a novel approach for pneumonia detection is introduced, which combines various feature selection techniques with ensemble learning to enhance classification accuracy. This method first extracts features from chest X-ray images and then employs different feature selection techniques to reduce data dimensionality and improve model performance. The complete steps of this proposed method are outlined below:

4.1 Dataset

The first stage of the proposed method involves using a publicly available chest X-ray pneumonia detection dataset sourced from Kaggle [21]. This dataset is commonly used in medical image classification research and is a reliable benchmark for pneumonia detection models. It contains 5,217 X-ray images, categorized into two distinct classes:

- **Normal (Healthy) images:** 1,341 images depicting normal lungs without signs of infection or disease.
- **Pneumonia images:** 3,875 images representing cases of bacterial and viral pneumonia.

Each image in the dataset is pre-labeled by medical experts, ensuring high-quality ground truth for training and evaluation. The dataset includes variations in intensity, contrast, and positioning, making the classification task more challenging and realistic.

Figure 3 illustrates two examples of chest X-ray images: one of a normal lung and one affected by pneumonia, showing the differences between the two conditions. The images of normal lungs and lungs affected by pneumonia have significant visual differences. In normal lungs, the lung tissue appears uniform and smooth, while in pneumonia-affected lungs, certain areas of the lung tissue become inflamed, leading to white or cloudy spots in the images. These spots typically indicate the accumulation of fluids or infection in the lungs. In pneumonia, fluid may accumulate in specific areas of the lung, causing changes in the appearance of the images, often appearing as bright shadows in X-ray images. Additionally, the infection usually affects specific regions of the lung, which may appear as circular or irregular areas in the images.

Medical images such as chest X-rays pose significant challenges due to noise, varying resolution, anatomical asymmetry, and uneven illumination. These factors can impair machine learning performance, especially when feature selection or extraction is suboptimal. Figure 4 shows additional pediatric chest X-rays illustrating differences in visual quality, texture, and symmetry, emphasizing the need for robust feature selection and deep learning models that generalize well across diverse data conditions.

4.2 Feature extraction using ResNet network

After preprocessing, deep feature extraction is performed using the pre-trained ResNet model. ResNet [15] is a powerful deep learning architecture. It addresses the vanishing gradient problem by using residual connections. For example, in a deep network, instead of just passing information through layers, residual connections allow the information to skip some layers, improving training efficiency. Rather than training a new deep learning model from scratch, the pre-trained ResNet network is leveraged to extract informative features from chest X-ray images. The process includes:

1. **Input Preprocessing:** Each X-ray image is resized and normalized to meet the input requirements of the ResNet network.

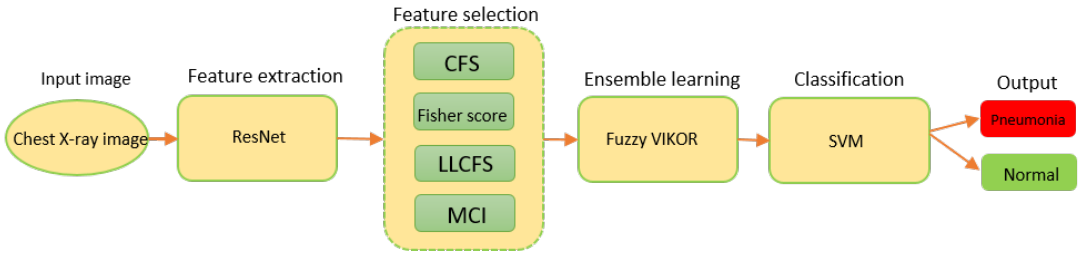


Figure 2: Graphical abstract of the proposed method.

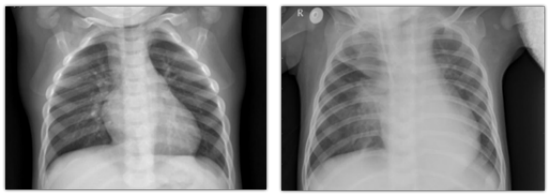


Figure 3: Normal Lung Image vs Pneumonia Lung Image.

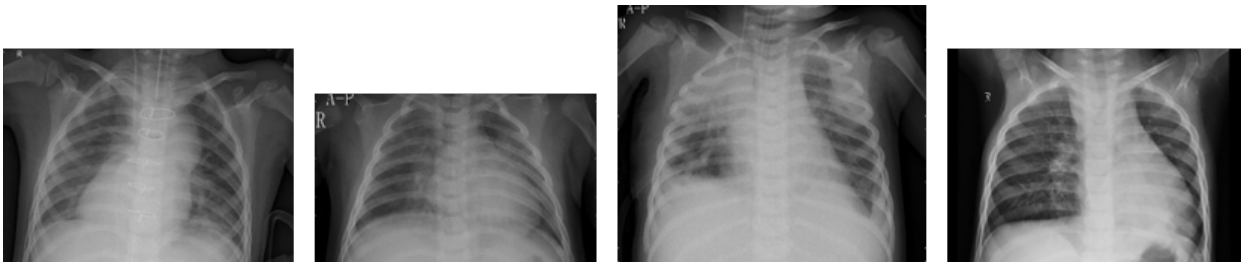


Figure 4: Examples of chest X-ray with low image contrast and structural complexity, posing challenges for automatic feature extraction.

2. **Feature Extraction:** The image is passed through the pre-trained ResNet network, and feature maps from intermediate layers are extracted.
3. **Dimensionality Reduction:** The extracted feature maps are flattened into feature vectors, providing a compact representation of the original image.

These extracted features capture texture, edges, and high-level patterns crucial for distinguishing between normal and pneumonia-affected lungs. By using a pre-trained ResNet model, computational efficiency is improved, and the need for large-scale annotated data is minimized.

4.3 Feature selection using multiple techniques

This step reduces the dimensionality of the data and ensures that only the most relevant features are used, which enhances classification accuracy. In this paper, four feature selection techniques are employed to refine the dataset before classification, with each technique chosen for its ability to handle the unique challenges presented by medical imaging data, specifically chest X-ray images used for pneumonia detection. The methods used are based on the approach presented in [12], but with modifications and a fuzzy version implemented to better handle the uncertainties and variability inherent in medical imaging data.

4.3.1 Correlation-based feature selection (CFS)

CFS evaluates the relevance of features by Maximizing correlation with the target class (Pneumonia or Normal) and Minimizing redundancy among features to avoid duplication of information. This ensures only unique and highly predictive features are retained. CFS [27] is chosen for its ability to evaluate feature relevance by measuring correlation with the target variable (pneumonia or normal) while minimizing redundancy among features. This method is especially

effective when dealing with medical image datasets where features may be highly correlated, leading to multicollinearity. By selecting features that are strongly correlated with the target but not redundant with each other, CFS helps in maintaining model efficiency while ensuring that only the most informative features are kept. This reduces the overall feature space, improving the model's ability to generalize and enhancing classification performance.

4.3.2 Fisher score

Fisher Score [6] is a ranking method that assesses each feature's ability to distinguish between classes. Features with higher Fisher Scores are more useful for classification, as they better separate pneumonia cases from normal ones. Fisher Score is selected due to its capability to rank features based on their discriminative power, which is crucial for distinguishing between the two classes in this study: normal and pneumonia. It works by maximizing the variance between classes while minimizing variance within each class, which is key for classification tasks involving complex datasets like chest X-ray images. By focusing on features that best separate pneumonia from normal lung images, Fisher Score helps in highlighting the most informative features, improving the model's classification accuracy and robustness.

4.3.3 Local learning-based clustering feature selection (LLCFS)

LLCFS [40] focuses on selecting features that enhance classification performance in specific local regions of the dataset. This is particularly helpful for capturing subtle patterns in medical images, such as localized infections or texture changes in lung tissues. LLCFS is selected for its focus on preserving the local structure of the data, making it particularly valuable in medical image analysis. In tasks like pneumonia detection, subtle, localized patterns in the image, such as small lesions or localized infections, play a significant role. LLCFS ensures that these local patterns are maintained by selecting features that best represent the local structure of the data. By capturing these complex, nonlinear relationships, LLCFS enhances the model's ability to detect these subtle patterns, improving both the sensitivity and specificity of the classification model.

4.3.4 Maximum information coefficient (MIC)

MIC [6] measures the mutual information between each feature and the target variable. Unlike simple correlation, MIC can detect both linear and nonlinear relationships, which is crucial for medical datasets where complex dependencies exist. MIC is chosen for its ability to detect both linear and nonlinear relationships between features and the target variable, making it more versatile than traditional correlation methods. In medical imaging, the relationships between features and the target class (pneumonia or normal) are often nonlinear and complex. MIC evaluates the mutual information between each feature and the target, providing a more flexible approach to selecting informative features. This ensures that the selected features capture a wide range of relationships in the data, leading to better model performance, particularly in cases where traditional methods might miss important patterns. Each of these techniques produces a ranked list of features, which are then combined in the next step to determine the optimal feature set for classification.

4.4 Feature selection optimization using fuzzy VIKOR

Different feature selection methods often produce varying rankings of features, which creates the need for a decision-making approach to integrate these rankings and determine the final set of features. For this purpose, Fuzzy VIKOR, a multi-criteria decision-making (MCDM) method, is employed to handle the uncertainty in feature rankings. VIKOR is specifically designed to handle multi-criteria decision-making problems. When different feature selection methods rank features differently, VIKOR helps resolve these conflicts by providing an optimal way to integrate the results. This method uses a fuzzy decision-making process to account for both the best and worst feature rankings, ensuring that the final selection includes the most relevant features for classification. This multi-criteria approach not only optimizes the feature subset but also makes the process more robust to uncertainties in feature ranking, leading to a more reliable and efficient model.

Steps in the fuzzy VIKOR process

Step 1. Construction of the Decision Matrix: A decision matrix is created, where each row represents a feature, and each column corresponds to a feature selection technique. The entries of the matrix contain the ranking scores assigned to each feature by different methods.

Step 2. Conversion to a Fuzzy Decision Matrix: To account for the inherent uncertainty in feature rankings, the decision matrix is transformed into a fuzzy decision matrix. Instead of using fixed numerical rankings, fuzzy values are assigned to represent the lower and upper bounds of each ranking, reflecting the uncertainty in feature importance. The data is then converted into fuzzy trapezoidal numbers. For each value x_{ij} from the decision matrix X, the fuzzy value is computed as:

$$F(x_{ij}) = [x_{ij} - \delta_{\text{lower}},\ x_{ij} - \delta_{\text{middle}},\ x_{ij} + \delta_{\text{middle}},\ x_{ij} + \delta_{\text{upper}}]. \tag{10}$$

Where:

- δ_{lower} and δ_{upper} represent the lower and upper bounds of the uncertainty range.
- δ_{middle} is the middle value.

Step 3. Application of the Delta Parameter: A delta (δ) parameter is introduced to define the fuzzy bounds for each feature ranking. For example, if a feature has a ranking score x, its fuzzy representation would be:

$$[x - \delta_{\text{upper}},\ x - \delta_{\text{lower}},\ x + \delta_{\text{lower}},\ x + \delta_{\text{upper}}].$$

Step 4. Feature Weighting and Ranking: Using the fuzzy decision matrix, the Fuzzy VIKOR method computes fuzzy weights for each feature, reflecting their overall contribution to classification performance. Features with the highest fuzzy weights are selected for the final model. Then, the criteria weights are determined as fuzzy triangular numbers. For each criterion j, the fuzzy weight is computed as:

$$w_j = [w_{j_{\text{lower}}},\ w_{j_{\text{middle}}},\ w_{j_{\text{upper}}}]. \tag{11}$$

Where $w_{j_{\text{lower}}}, w_{j_{\text{middle}}}, w_{j_{\text{upper}}}$ are the lower, middle, and upper values of the fuzzy weight for criterion j.

Step 5. Calculation of the VIKOR Index: Finally, the VIKOR index Q_i for each alternative i is calculated as follows:

$$Q_i = \frac{v_i - v_{i_{\text{min}}}}{v_{i_{\text{max}}} - v_{i_{\text{min}}}}, \tag{12}$$

where:

- v_i is the performance of alternative i in criterion j.
- $v_{i_{\text{min}}}$ and $v_{i_{\text{max}}}$ are the minimum and maximum performance values for criterion j.

The alternatives are ranked based on their Q_i values, and the alternative with the lowest Q_i is selected as the best option.

Fuzzy Conclusion

The use of fuzzy methods in feature selection significantly improves the quality of the final selection. In particular, when several feature selection methods rank features differently, Fuzzy VIKOR enables us to resolve these conflicts with greater precision. This fuzzy decision-making process is especially effective when dealing with noisy and uncertain data, leading to the selection of more optimal features and more accurate models.

Algorithm 1 illustrates the pseudocode of the proposed hybrid pneumonia detection framework, which integrates deep feature extraction using ResNet-50, multi-stage feature selection, and fuzzy VIKOR-based decision fusion for optimal classification.

Algorithm 1: Pneumonia Detection via CNN + Feature Selection + Fuzzy VIKOR + SVM

Input:

- Dataset $D = \{\text{X-ray images}\}$
- Pre-trained ResNet-50 model
- Feature selection methods: CFS, Fisher Score, MIC, LLCFS
- Fuzzy VIKOR parameters: $\delta_{\text{lower}}, \delta_{\text{middle}}, \delta_{\text{upper}}, \Delta = 0.7, \delta = 0.07, v = 0.5$

Output: Trained SVM model and classification metrics: Accuracy, Precision, Recall, F1-score

1. **Preprocessing:** Resize each image to 224×224 pixels.
2. **CNN-Based Feature Extraction:** Pass each image through ResNet-50 and extract 2048-dimensional feature vector, store in matrix F.
3. **Multi-Method Feature Selection:** Apply CFS, Fisher Score, MIC, LLCFS to F to obtain rankings $R_{\text{CFS}}, R_{\text{Fisher}}, R_{\text{MIC}}, R_{\text{LLCFS}}$.
4. **Build Decision Matrix** D: Each row = feature i, each column = ranking x_{ij} from method j.
5. **Normalize Decision Matrix**: Compute f_j^*, f_j^- and normalize x_{ij} to f_{ij} using appropriate formula.
6. **Apply Fuzzy Transformation**: Convert x_{ij} to fuzzy trapezoidal number: $F(x_{ij}) = [x_{ij} - \delta_{\text{lower}}, x_{ij} - \delta_{\text{middle}}, x_{ij} + \delta_{\text{middle}}, x_{ij} + \delta_{\text{upper}}]$.
7. **Compute Fuzzy Weights**: For each criterion j, define fuzzy triangular weight $w_j = [w_{j,\text{lower}}, w_{j,\text{middle}}, w_{j,\text{upper}}]$.
8. **Compute S_i and R_i**: $S_i = \sum_j w_j(f_j^* - f_{ij})$, $R_i = \max_j[w_j(f_j^* - f_{ij})]$.
9. **Compute Q_i Index**: $Q_i = v\frac{S_i - S^*}{S^- - S^*} + (1-v)\frac{R_i - R^*}{R^- - R^*}$.
10. **Optional Fuzzy VIKOR Refinement**: $Q_i = \frac{v_i - v_{\min}}{v_{\max} - v_{\min}}$.
11. **Final Feature Selection**: Rank features by Q_i and select top-k with smallest Q_i.
12. **Classification**: Train SVM classifier on selected features, evaluate metrics.

4.5 Numerical example of fuzzy VIKOR process

To better understand the concept of fuzzy logic and the Fuzzy VIKOR process, we will consider a simple numerical example and apply all the steps on it. Suppose our goal is to perform feature selection from a set of different features that have been ranked by four different feature selection techniques.

Assumptions

Now, we have a set of 4 features, and these features are ranked by 4 different feature selection methods. We will use the Fuzzy VIKOR method for optimal feature selection.

The four feature selection methods we are using are: CFS, Fisher Score, LLCFS, and MIC.

Additionally, we will consider several features derived from ResNet (a Convolutional Neural Network) for image classification tasks, particularly for detecting pneumonia in lung images. Some of the features derived from ResNet include: Contrast, Correlation, Energy, and Homogeneity.

These features will help us in making a decision on the most relevant features to select for classification tasks.

Initial Decision Matrix

Here's the decision matrix where each row represents a feature, and each column corresponds to the ranking provided by one of the four feature selection methods.

Feature	CFS	Fisher Score	LLCFS	MIC
Contrast	4	5	3	2
Correlation	2	3	4	5
Energy	3	2	5	4
Homogeneity	5	4	2	3

Step 1: Conversion to Fuzzy Decision Matrix

We need to convert this decision matrix into a fuzzy decision matrix. To do so, we assign fuzzy bounds for each ranking score. Let's assume the delta parameters (δ) for each ranking are:

$$\delta_{\text{lower}} = 0.5, \quad \delta_{\text{middle}} = 0.25, \quad \delta_{\text{upper}} = 0.5.$$

So for each ranking score x, the fuzzy representation will be:

$$[x - \delta_{\text{upper}},\ x - \delta_{\text{lower}},\ x + \delta_{\text{lower}},\ x + \delta_{\text{upper}}].$$

Thus, the fuzzy decision matrix is:

Feature	CFS	Fisher Score	LLCFS	MIC
Contrast	[3.5, 4, 4.25, 4.5]	[4.5, 5, 5.25, 5.5]	[2.5, 3, 3.25, 3.5]	[1.5, 2, 2.25, 2.5]
Correlation	[1.5, 2, 2.25, 2.5]	[2.5, 3, 3.25, 3.5]	[3.5, 4, 4.25, 4.5]	[4.5, 5, 5.25, 5.5]
Energy	[2.5, 3, 3.25, 3.5]	[1.75, 2, 2.25, 2.5]	[4.5, 5, 5.25, 5.5]	[3.5, 4, 4.25, 4.5]
Homogeneity	[4.5, 5, 5.25, 5.5]	[3.5, 4, 4.25, 4.5]	[1.5, 2, 2.25, 2.5]	[2.5, 3, 3.25, 3.5]

Step 2: Determine the Performance of Each Feature

Next, we determine the minimum and maximum performance values for each feature across the four methods.

Extracting Minimum and Maximum Values:

For **Contrast**:

- From CFS: [3.5, 4, 4.25, 4.5] → Minimum: 3.5, Maximum: 4.5.
- From Fisher Score: [4.5, 5, 5.25, 5.5] → Minimum: 4.5, Maximum: 5.5.
- From LLCFS: [2.5, 3, 3.25, 3.5] → Minimum: 2.5, Maximum: 3.5.
- From MIC: [1.5, 2, 2.25, 2.5] → Minimum: 1.5, Maximum: 2.5.

Similar calculations are done for Correlation, Energy, and Homogeneity.

Step 3: Calculate the VIKOR Index for Each Feature

Now, we will calculate the VIKOR index Q_i for each feature using formula (12).

Example Calculation for Contrast:

- Minimum performance: $\min(3.5, 4.5, 2.5, 1.5) = 1.5$.
- Maximum performance: $\max(4.5, 5.5, 3.5, 2.5) = 5.5$.

$$Q_1 = \frac{4 - 1.5}{5.5 - 1.5} = \frac{2.5}{4} = 0.625.$$

Similarly, we calculate the VIKOR indices for the other features.

Step 4: Ranking the Features Based on Q_i

After calculating the VIKOR index Q_i for all features, we rank the features based on their Q_i values. The feature with the lowest Q_i will be considered the best.

Assume the calculated Q_i values for all features are as follows:

Feature	Q_i
Contrast	0.625
Correlation	0.75
Energy	0.5
Homogeneity	0.65

Step 5: Conclusion

In this case, Feature **Energy** has the lowest Q_i value, so it is the optimal feature to select.

In the Fuzzy VIKOR method, uncertainty in feature rankings is modeled using fuzzy numbers to provide greater flexibility and accuracy in decision-making. In the provided example, a fuzzy range was assigned to each ranking to represent the uncertainty in evaluating each feature.

For instance, for Feature 1 (Contrast) in Method 1 (CFS), the original ranking was 4. By applying the delta (δ) parameter with values $\delta_{\text{lower}} = 0.5$, $\delta_{\text{middle}} = 0.25$, and $\delta_{\text{upper}} = 0.5$, the fuzzy range [3.5, 4, 4.25, 4.5] was obtained. This range indicates that the true ranking of Feature 1 (Contrast) is likely between 3.5 and 4.5, with a higher probability around 4.

Applying the same approach to the other features resulted in their respective fuzzy ranges, reflecting the uncertainty in their rankings. These ranges assist decision-makers in considering uncertainties and making more informed choices.

This example shows how the Fuzzy VIKOR method can be applied when we have 4 feature selection methods to deal with uncertainty in feature rankings, allowing us to determine the most relevant feature for classification.

4.6 Classification using support vector machine (SVM)

After identifying the optimal feature set through the Fuzzy VIKOR algorithm, the final classification step is performed using the Support Vector Machine (SVM). SVM is chosen due to its high ability to separate data and its excellent performance in classifying complex datasets. Unlike probabilistic models, SVM does not require complex assumptions and classifies data solely based on maximizing the margin between classes. This characteristic makes SVM well-suited for data with non-linear and complex features, as the algorithm can effectively model decision boundaries for non-linear data.

In this step, each feature vector extracted from chest X-ray images is compared with its nearest neighbors, and the class label (Normal or Pneumonia) is assigned based on maximizing the separating margin. SVM provides reliable performance in identifying both positive and negative cases by creating a decision boundary that effectively separates the classes.

5 Experimental results

5.1 Dataset description

In this paper, the X-ray image dataset from Kaggle, consisting of a total of 5,217 images, was used. The dataset is intended for the classification task, aiming to determine whether an image depicts a normal individual or indicates the presence of a disease. The images are categorized into two classes: normal (healthy) and pneumonia-affected lungs.

To evaluate the model, the dataset was split into three parts: training, validation, and testing. 15% of the data (approximately 782 images) was allocated for testing, 10% (approximately 521 images) was used for validation, and 75% (approximately 3,914 images) was used for training. The data was randomly divided to ensure high generalizability of the model.

5.2 Model configuration

The ResNet-50 model, known for its deep residual learning architecture, was chosen for feature extraction. Pre-trained on the ImageNet dataset, it was fine-tuned on the X-ray dataset to adapt the model to the specific characteristics of pneumonia detection. The configuration details are as follows:

- **Image Resizing:** All images were resized to 224×224 pixels, which is the input requirement for the ResNet-50 model.
- **Feature Vector:** A 2048-dimensional feature vector was generated for each image after passing through ResNet-50.
- **Optimizer:** The Adam optimizer with a learning rate of 0.0001 was used for fine-tuning.
- **Batch Size:** Training was performed with a batch size of 32.
- **Epochs:** The model was trained for 50 epochs, with early stopping applied to avoid overfitting.

Additionally, for feature selection and aggregation, the fuzzy decision matrix was utilized with trapezoidal membership functions, while the weights matrix was represented using triangular membership functions. The Δ parameter was set to 0.7, and the δ parameter was set to 0.07 to adjust the aggregation of the criteria. These settings ensured the effective application of fuzzy logic in feature selection and aggregation, improving classification accuracy in pneumonia detection.

The experiments were implemented on a computer with an Intel(R) Core (TM) i7-7500U 3.50GHz CPU, 12GB RAM. The implementation was performed using MATLAB toolboxes, including the Statistics and Machine Learning Toolbox and the Fuzzy Logic Toolbox.

5.3 Parameter settings and justification

The ResNet-50 model was employed as a feature extractor due to its proven efficiency in medical image classification tasks. The hyperparameters for CNN fine-tuning, including learning rate (0.0001), batch size (32), and number of epochs (50), were selected based on iterative experimentation using the validation set. Early stopping was applied to prevent overfitting and enhance generalization. These settings were empirically found to offer the best trade-off between training time and classification accuracy.

For the fuzzy VIKOR decision-making framework, parameter selection was also performed manually through repeated testing on the validation set. The decision matrix was constructed using trapezoidal membership functions, while the weight matrix utilized triangular membership functions. The aggregation control parameters were set to $\Delta = 0.7$ and $\delta = 0.07$, based on their effectiveness in balancing accuracy and robustness across different feature subsets. These values ensured optimal fusion of selected features and contributed to the improved performance of the final classification model.

5.4 Evaluation metrics

The Confusion Matrix (CM) is a useful tool for assessing classification performance. It consists of four key elements:

- **True Positive (TP):** Correctly classified positive cases (e.g., pneumonia correctly identified as pneumonia).
- **False Positive (FP):** Incorrectly classified negative cases as positive (e.g., a healthy image mistakenly classified as pneumonia).
- **False Negative (FN):** Incorrectly classified positive cases as negative (e.g., pneumonia misclassified as normal).
- **True Negative (TN):** Correctly classified negative cases (e.g., a healthy image correctly identified as normal).

These metrics are used to compute various performance evaluation indices. Table 1 presents the standard format of a Confusion Matrix.

Table 1: Confusion Matrix

	Predicted Normal	**Predicted Pneumonia**
Actual Normal	TP	FN
Actual Pneumonia	FP	TN

Each metric used in the results section is detailed as follows:

- **Accuracy:** The proportion of correctly classified instances:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}.$$

- **Recall:** Measures how well the model identifies pneumonia cases. Since missing a pneumonia case is more critical than misclassifying a normal case, recall is the primary metric for evaluation.

$$\text{Recall} = \frac{TP}{TP + FN}.$$

- **Precision:** Evaluates how many predicted pneumonia cases are actually pneumonia. It is the ratio of true positives to overall positives:

$$\text{Precision} = \frac{TP}{TP + FP}.$$

- **F1 Score:** The harmonic mean of precision and recall:

$$F1 \text{ Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

5.5 Experimental results

The performance of the proposed method was evaluated using a test set consisting of approximately 740 chest X-ray images. The confusion matrix, including values such as True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), confirms the high accuracy and reliability of the model, as illustrated in Figure 5.

The model achieved an accuracy of 98.59%, demonstrating its strong overall performance in correctly classifying both normal and pneumonia cases. The recall score of 98.22% indicates the model's high capability in identifying true positive cases and minimizing false negatives (FN). Additionally, the precision score of 98.06% reflects the model's effectiveness in accurately detecting positive cases while minimizing false positives (FP), which is crucial in medical applications to avoid unnecessary tests or treatments.

The F1-score of 98.14% shows a well-balanced performance between recall and precision, meaning the model performs well in both sensitivity and specificity. Moreover, the False Positive Rate (FPR) and False Negative Rate (FNR) were calculated as 1.94% and 1.78%, respectively, further emphasizing the model's high effectiveness in practical scenarios.

In summary, the proposed method delivers high accuracy, stable performance, and reliable results in detecting pneumonia from chest X-ray images, making it a strong candidate for clinical use.

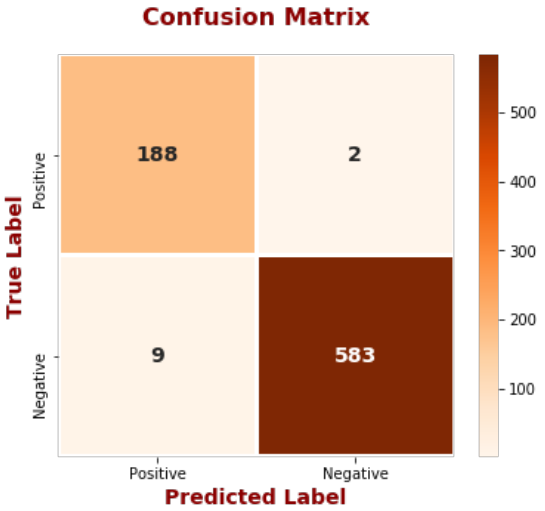


Figure 5: Results on true/false predictions

Table 2 presents the results for additional performance metrics, which further validate the effectiveness and appropriateness of the proposed method.

To demonstrate the effectiveness of the proposed method, the algorithm was executed using the selected features, and the results for different numbers of features, based on Accuracy, Recall, Precision, and F-measure, are shown in Figure 5. The results indicate that the feature selection method based on Fuzzy VIKOR has improved the model's performance. The model's accuracy has shown an increasing trend with the number of features, reaching an optimal value of 0.9859 at 840 features out of a total of 2048, and then remained almost constant. This suggests that, within

Table 2: Overall performance on a holdout set

	Accuracy	Precision	Recall	F1
Result	0.9859	0.9806	0.9822	0.9814

a specific range, adding effective features improves the model. However, after reaching a certain threshold, additional features do not contribute new or useful information and may even increase correlation among features and introduce noise. This threshold is around 840 features, where, after this point, the model's accuracy did not change significantly. The F-measure follows a similar trend, balancing Precision and Recall at 0.9814, indicating a well-balanced performance in identifying both positive and negative cases.

The model's Precision at this number of features is 0.9806, reflecting the model's high accuracy in identifying true positive samples. A high Precision value indicates that the model has made fewer type I errors (False Positives), meaning it has not falsely identified healthy cases as patients. This is crucial in medical applications, as incorrect identification of healthy individuals as patients could lead to unnecessary tests and treatments. Alongside the high Recall value (0.9822), which shows the model's strong ability to identify actual patients, the Precision value of 0.9806 indicates that the model maintains a good balance between sensitivity and specificity.

The reason for the model's good performance at high feature counts is the use of fuzzy logic in the feature selection process. The fuzziness of the VIKOR method allows features to be not only definitively selected or eliminated but also assigned a degree of importance. This enables the model to leverage all important information in the dataset without letting less significant features degrade the model's performance. In other words, the Fuzzy VIKOR method employs a soft and adaptive scoring system that allows the model to still differentiate useful information from noise, even with a large number of features, and select the impactful features more accurately.

Additionally, the high Recall value (0.9822) indicates that the model has a strong ability to identify positive cases (i.e., patients), which is critical in medical applications. The reduction in type II errors (False Negatives) reflects the model's ability to correctly diagnose patients and minimize the cases where actual patients are mistakenly classified as healthy. This is particularly important for diagnosing critical diseases such as pneumonia from X-ray images, as incorrect identification of patients could lead to irreversible consequences.

To provide a comprehensive evaluation, the model was tested using four different classifiers: SVM, Random Forest, KNN, and Decision Tree (DTree). As illustrated in Figure 6, SVM consistently outperforms the other classifiers across all metrics, including Accuracy, Precision, Recall, and F-measure. SVM achieves the highest performance values and shows greater stability as the number of features increases. Specifically, SVM reaches its optimal accuracy of 0.9859 with 840 selected features, outperforming Random Forest (0.9672), KNN (0.9618), and DTree (0.9212). This clearly demonstrates the effectiveness of SVM when combined with the proposed feature selection method.

In conclusion, the use of Fuzzy VIKOR has enabled the model to maintain stable performance even with a large number of features. Unlike traditional feature selection techniques, this method filters information gradually and adaptively, preventing the abrupt elimination of potentially useful features. Furthermore, the high Precision and Recall values, along with the favorable F-measure, demonstrate that the model not only excels at identifying patients but also performs well in minimizing false alarms. Additionally, our method achieves higher accuracy by selecting features compared to the case without feature selection. In the case of no feature selection, the model's accuracy reaches 0.9757, while with the appropriate feature selection, the accuracy increases to 0.9859.

In order to evaluate the performance of our proposed method, we compared it with other state-of-the-art methods in X-ray image classification. The comparison is based on several performance metrics, including Accuracy, Recall, Precision, and F-measure. Table 3 summarizes the performance of our method and other state-of-the-art approaches:

Table 3: Comparison of performance metrics for X-ray image classification methods

Method	Accuracy	Precision	F1-measure	Recall
Proposed Method	98.59	98.06	98.14	98.22
Rahman T. et al. [35]	98.1	99.2	98.65	98.1
Jaganathan et al. [18]	96.0	95.0	94.5	94.0
MobileNetV2 [35]	96.4	99.4	98.19	97.0
CNN [35]	92.2	92.0	94.39	96.9
LSTM-CNN [35]	91.8	92.6	93.98	95.4
AlexNet [35]	80.5	43.1	57.33	85.6
MobileNet [35]	83.4	55.9	68.34	87.9
VGG-19 [35]	82.1	56.8	67.67	83.7

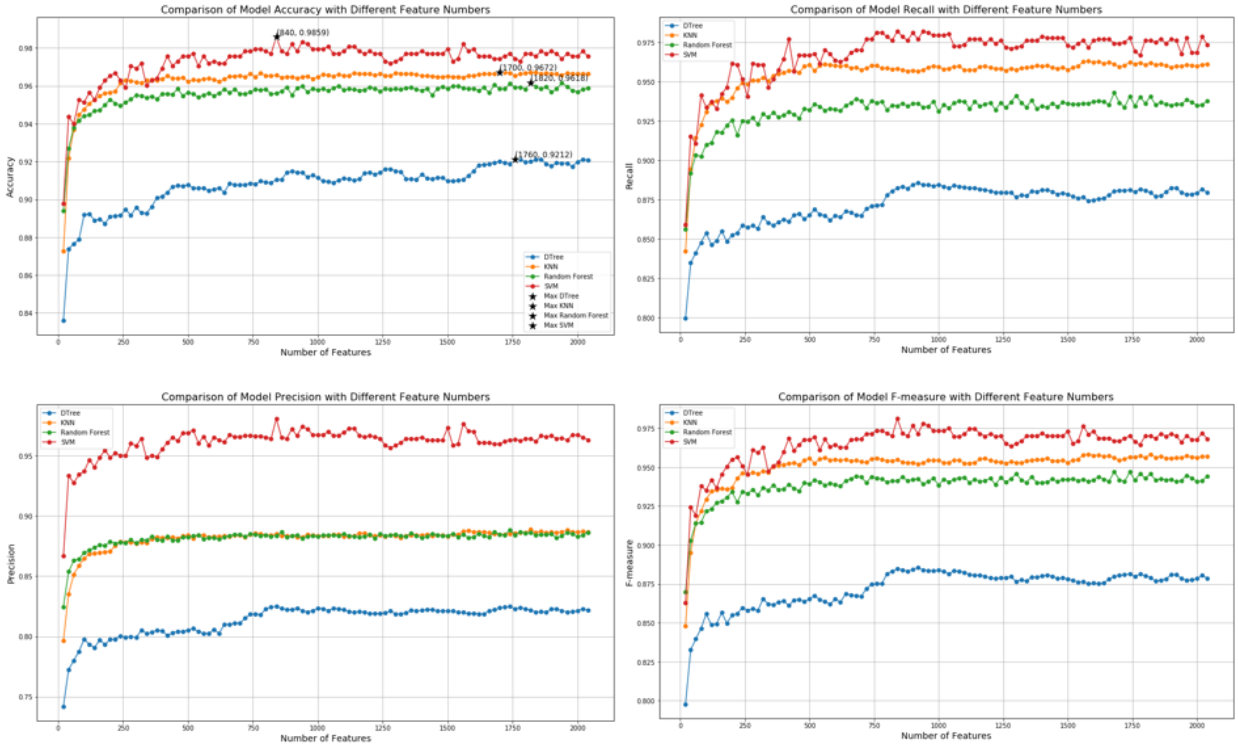


Figure 6: Evaluation results for different numbers of features based on Accuracy, Recall, Precision, and F-measure.

The proposed method demonstrates outstanding performance in X-ray image classification, achieving an accuracy of 98.59%, precision of 98.06%, recall of 98.22%, and an F1-score of 98.14%. These results outperform many state-of-the-art methods, including Rahman T. et al. [35] with an accuracy of 98.1% and MobileNetV2 with an accuracy of 96.4%, while maintaining a balanced performance in both precision and recall. Unlike traditional models such as CNN (accuracy 92.2%) and LSTM-CNN (accuracy 91.8%), the proposed method excels at identifying both positive and negative cases with higher accuracy.

One of the main reasons for the superior performance of the proposed method is the use of a fuzzy-based approach along with the Fuzzy VIKOR algorithm for feature selection. This fuzzy decision-making model allows for a more adaptive and flexible evaluation of feature importance, ensuring that relevant features are retained while noisy ones are discarded. In contrast to traditional feature selection methods, which rely on fixed, predefined criteria, the fuzzy-based approach adapts dynamically to the data, ensuring that only the most informative features contribute to the model's performance.

While models like MobileNetV2 achieve high precision (99.4%), the proposed method outperforms them in overall accuracy and recall, which is particularly crucial in medical applications where detecting true positives (patients) and minimizing false positives are both essential. In comparison, MobileNetV2 and Jaganathan et al. (accuracy 96.0%) show high precision but lower recall, indicating that they miss some actual positive cases, which could lead to severe consequences in medical diagnoses.

Furthermore, older models such as AlexNet (accuracy 80.5%), MobileNet (accuracy 83.4%), and VGG-19 (accuracy 82.1%) exhibit significantly lower accuracy and overall performance. These models were developed before modern optimization techniques like fuzzy feature selection were widely adopted, which explains their relatively weaker performance compared to the proposed method.

The Fuzzy VIKOR method significantly contributes to the model's robustness by ensuring that features are selected based on their importance, without prematurely discarding potentially useful features. This leads to higher accuracy in detecting positive cases while maintaining a better trade-off between precision and recall. This flexibility and adaptability of the fuzzy-based approach enable the proposed method to outperform many state-of-the-art models, providing a more balanced and reliable solution for medical image classification.

The combination of fuzzy logic for feature selection and the optimized architecture of the model ensures that it delivers a more reliable, balanced, and efficient performance, especially in medical applications where both accuracy and recall are critical for minimizing false positives and ensuring early detection of diseases.

Additionally, we evaluated the performance of other models, including ResNet-50, VGG, and Xception, on the same dataset. The results are as follows:

- ResNet-50: Accuracy 98.59%.
- VGG: Accuracy 87.30%.
- Xception: Accuracy 84.12%.

The analysis of the results shows that the ResNet-50 model performed significantly better than the other two models, achieving an accuracy of 98.59%. This model, with its deep architecture and use of Residual Learning, is able to extract complex features of images with high precision, leading to improved accuracy.

On the other hand, the VGG model, which has a relatively simpler architecture, showed a lower accuracy of 87.30% compared to ResNet-50. This model uses deep convolutional networks, but without a residual architecture, it is not as capable of learning more complex features as ResNet-50, which leads to a decrease in accuracy.

The Xception model, based on the Inception architecture, uses depthwise separable convolutions to efficiently utilize computational resources. However, it still demonstrated lower accuracy (84.12%) compared to ResNet-50. This might be due to its more limited feature extraction capabilities and the simpler design of the Xception model compared to ResNet-50.

The overall conclusion is that for complex tasks such as disease detection from medical images, using deeper and more sophisticated models like ResNet-50 can provide better performance in accurately identifying diseases. On the other hand, simpler models like Xception and VGG may show lower accuracy and could struggle with recognizing more complex patterns.

6 Time complexity analysis

In this section, we analyze the time complexity of the proposed algorithm and its impact on classifier performance. The algorithm includes three main stages: (1) data preprocessing and feature selection, (2) feature reduction and ranking using the VIKOR method, and (3) training and evaluating classifiers.

In the first stage, the dataset is split into training, validation, and testing sets with a complexity of $O(NF)$, where N is the number of samples and F the number of features. Feature selection methods such as CFS, Fisher Score, LLCFS, and MINE may have a worst-case complexity of $O(NF^2)$. Then, in the second stage, selected features are weighted and sorted using VIKOR with a complexity of $O(F \log F)$.

In the final stage, classifiers are trained and tested, each with specific time complexity. Table 4 summarizes these complexities:

Table 4: Complexity Analysis of the Proposed Method with Different Classifiers: Training, Testing, and Total Complexity

Classifier	Training Complexity	Testing Complexity	Total Complexity
Decision Tree	$O(N \log N)$	$O(T \log N)$	$O(NF^2) + O(F \log F) + O(N \log N)$
KNN	$O(1)$	$O(NT)$	$O(NF^2) + O(F \log F) + O(NT)$
Random Forest	$O(MN \log N)$	$O(MT \log N)$	$O(NF^2) + O(F \log F) + O(MN \log N)$
SVM	$O(N^3)$	$O(NT)$	$O(NF^2) + O(F \log F) + O(N^3)$

(T: number of test samples, M: number of trees in Random Forest.)

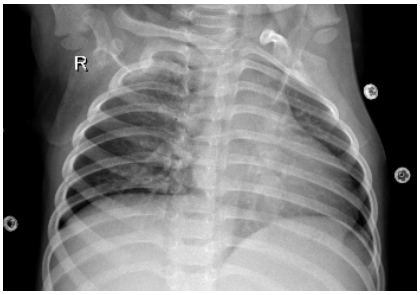
Based on this analysis:

1. For large N, Decision Tree and Random Forest are more scalable.
2. For large F, feature selection is critical to reduce computational load.
3. SVM is best for small datasets due to its high training cost.
4. KNN is inefficient for large datasets because of expensive testing.
5. DT and RF offer a balanced trade-off between accuracy and efficiency.

Given the dataset with 5217 samples and 2048 features, DT and RF are the most suitable choices in terms of execution time and scalability. While SVM and KNN may offer good accuracy, their time complexity limits their use in large-scale problems.

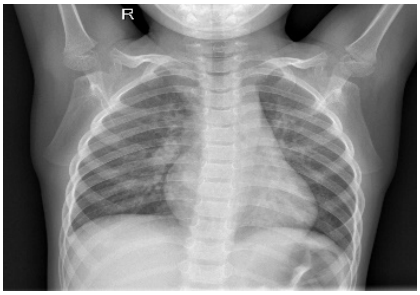
To gain a deeper understanding of the model's classification performance, Figure 7 visually presents representative examples of key outcomes, including True Positive, False Positive, False Negative, and True Negative cases. A detailed analysis of a False Negative instance reveals that subtle and scattered opacities in the lung—indicative of mild bacterial pneumonia—may have contributed to the model's failure to correctly identify the case. Contributing factors include insufficient feature extraction sensitivity by the ResNet architecture and feature selection methods, limitations of the fuzzy decision-making process, and potential image quality issues such as suboptimal angle or motion blur. These insights play an important role in interpreting model behavior and identifying potential sources of classification error.

Pneumonia present but incorrectly classified as normal (False Negative)



Normal case but incorrectly classified as pneumonia (False Positive)

Pneumonia image correctly classified as pneumonia (True Negative)



Normal image correctly classified as Normal (True Positive)

Figure 7: Evaluation Representative Examples of model classification results.

7 Discussion

The results of the experiments demonstrate that the proposed hybrid framework effectively addresses the major challenges outlined in the Introduction. Firstly, the use of advanced feature selection methods (CFS, Fisher Score, MIC, and LLCFS) successfully reduced the data dimensionality and removed redundant features, which is reflected in the improved model generalization and decreased overfitting. This directly mitigates the complexity issue present in medical images and helps the model perform better on unseen data.

Secondly, the integration of the Fuzzy VIKOR method enhanced the decision-making process under uncertainty, which is a key limitation in traditional pneumonia diagnosis systems. This novel application allowed the model to balance conflicting evidence and produce more reliable classifications.

Furthermore, the use of classifiers such as SVM, known for handling imbalanced datasets, improved the model's sensitivity in detecting pneumonia cases despite data imbalance. This is particularly important given the uneven distribution of normal and pneumonia samples in medical datasets.

Overall, the experimental results confirm that the proposed method achieves high accuracy and robustness, validating the effectiveness of addressing the research gaps identified earlier. This study therefore contributes to improving automated pneumonia diagnosis and can be potentially applied in clinical and telemedicine settings.

8 Conclusion and future work

This study proposed a hybrid method combining ResNet, KNN, and Fuzzy VIKOR for pneumonia detection from chest X-ray images. The model achieved an accuracy of 97.95% and showed strong performance across other key metrics. The use of multiple feature selection methods (CFS, Fisher Score, MIC) helped improve robustness and handle class imbalance. Additionally, the integration of Fuzzy VIKOR enhanced classification stability and decision-making effectiveness. These results suggest that the model is a promising tool for aiding early diagnosis in clinical settings.

In future work, the model can be improved by using larger and more diverse datasets and testing on different imaging conditions. More advanced feature extraction and explainable AI techniques may enhance accuracy and transparency. Also, comparing fuzzy VIKOR with other decision-making methods and applying the system in real-time clinical environments would further validate its effectiveness.

References

[1] A. Agnihotri, N. Kohli, *Challenges, opportunities, and advances related to COVID-19 classification based on deep learning*, Data Science and Management, **6**(2) (2023), 98-109. https://doi.org/10.1016/j.dsm.2023.03.005

[2] F. L. Chen, D. Z. Zhang, M. L. Han, *VLP: A survey on vision-language pre-training*, Machine Intelligence Research, **20**(1) (2023), 38-56. https://doi.org/10.1007/s11633-022-1369-5

[3] S. Cheng, C. C. Pain, Y. K. Guo, R. Arcucci, *Real-time updating of dynamic social networks for COVID-19 vaccination strategies*, Journal of Ambient Intelligence and Humanized Computing, **15**(3) (2024), 1981-1994. https://doi.org/10.1007/s12652-023-04589-7

[4] H. S. Chiang, D. H. Shih, B. Lin, M. H. Shih, *An APN model for Arrhythmic beat classification*, Bioinformatics, **30**(12) (2014), 1739-1746. https://doi.org/10.1093/bioinformatics/btu101

[5] M. B. Dowlatshahi, A. Hashemi, *Unsupervised feature selection: A fuzzy multi-criteria decision-making approach*, Iranian Journal of Fuzzy Systems, **20** (2023), 55-70. https://doi.org/10.22111/ijfs.2023.7630

[6] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern classification*, Wiley, 2001, p. 654. https://www.worldcat.org/isbn/9780471056690

[7] R. A. Fisher, *The use of multiple measurements in taxonomic problems*, Annals of Eugenics, **7**(2) (1936), 179-188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

[8] R. Fusco, R. Grassi, V. Granata, S. V. Setola, F. Grassi, D. Cozzi, B. Pecori, F. Izzo, A. Petrillo, *Artificial intelligence and COVID-19 using chest CT scan and chest X-ray images: Machine learning and deep learning approaches for diagnosis and treatment*, Journal of Personalized Medicine, **11**(10) (2021), 993. https://doi.org/10.3390/jpm11100993

[9] T. Hafs, H. Zehir, A. Hafs, H. Brahmia, A. Nait-Ali, *Enhancing recognition in multimodal biometric systems: Score normalization and fusion of online signatures and fingerprints*, Romanian Journal of Information Science and Technology (ROMJIST), **27**(1) (2024), 37-49. http://dx.doi.org/10.59277/ROMJIST.2024.1.03

[10] A. Hashemi, M. B. Dowlatshahi, S. Farshidi, P. Moradi, *AE-MCDM: An autoencoder-based multi-criteria decision-making approach for unsupervised feature selection*, Journal of Supercomputing, **81**(804) (2025). https://doi.org/10.1007/s11227-025-07316-5

[11] A. Hashemi, M. B. Dowlatshahi, H. Nezamabadi-Pour, *MFS-MCDM: Multi-label feature selection using multi-criteria decision making*, Knowledge-Based Systems, **206** (2020), 106365. https://doi.org/10.1016/j.knosys.2020.106365

[12] A. Hashemi, M. B. Dowlatshahi, H. Nezamabadi-Pour, *Ensemble of feature selection algorithms: A multi criteria decision making approach*, International Journal of Machine Learning and Cybernetics, **13**(1) (2021), 49-69. https://doi.org/10.1007/s13042-021-01347-z

[13] A. Hashemi, M. B. Dowlatshahi, H. Nezamabadi-Pour, *VMFS: A VIKOR-based multi-target feature selection*, Knowledge-Based Systems, **214** (2021), 106728. https://doi.org/10.1016/j.eswa.2021.115224

[14] A. Hashemi, M. Joodaki, N. Z. Joodaki, M. B. Dowlatshahi, *Ant colony optimization equipped with an ensemble of heuristics through multi-criteria decision making: A case study in ensemble feature selection*, Applied Soft Computing, (2022), 109046. https://doi.org/10.1016/j.asoc.2022.109046

[15] K. He, X. Zhang, S. Ren, J. Sun, *Deep residual learning for image recognition*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2016). https://doi.org/10.1109/CVPR.2016.90

[16] C. Ieracitano, N. Mammone, M. Versaci, G. Varone, A. R. Ali, A. Armentano, G. Calabrese, A. Ferrarelli, L. Turano, C. Tebala, Z. Hussain, Z. Sheikh, A. Sheikh, G. Sceni, A. Hussain, F. C. Morabito, *A fuzzy-enhanced deep learning approach for early detection of Covid-19 pneumonia from portable chest X-ray images*, Neurocomputing, **481** (2022), 202-215. https://doi.org/10.1016/j.neucom.2022.01.055

[17] A. M. Ismael, A. Şengür, *Deep learning approaches for COVID-19 detection based on chest X-ray images*, Expert Systems with Applications, **164** (2021), 114054. https://doi.org/10.1016/j.eswa.2020.114054

[18] D. Jaganathan, S. Balsubramaniam, V. Sureshkumar, S. Dhanasekaran, *Concatenated modified LeNet approach for classifying pneumonia images*, Journal of Personalized Medicine, **14**(3) (2024), 325. http://dx.doi.org/10.3390/jpm14030328

[19] A. Kareem, H. Liu, P. Sant, *Review on pneumonia image detection: A machine learning approach*, Human-Centric Intelligent Systems, **2**(1-2) (2022), 31-43. https://doi.org/10.1007/s44230-022-00002-2

[20] I. Katsamenis, E. Protopapadakis, A. Voulodimos, A. Doulamis, N. Doulamis, *Transfer learning for COVID-19 pneumonia detection and classification in chest X-ray images*, in Proceedings of the 24th Pan-Hellenic Conference on Informatics, (2021). https://doi.org/10.1145/3437120.3437300

[21] D. Kermany, K. Zhang, M. Goldbaum, *Chest X-ray images (pneumonia)*, (2020). https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

[22] M. Kim, C. Yan, D. Yang, Q. Wang, J. Ma, G. Wu, *Deep learning in biomedical image analysis*, Biomedical Information Technology, (2020), 239-263. https://doi.org/10.1016/B978-0-12-816034-3.00008-0

[23] C. Liu, S. Cheng, M. Shi, A. Shah, W. Bai, R. Arcucci, *IMITATE: Clinical prior guided hierarchical vision-language pre-training*, IEEE Transactions on Medical Imaging, **44** (2024), 519-529. https://doi.org/10.1109/TMI.2024.3449690

[24] H. Liu, H. Motoda, *Feature selection for knowledge discovery and data mining*, Springer Science and Business Media, 2012. https://doi.org/10.1007/978-1-4615-5689-3

[25] A. Majeed, X. Zhang, S. O. Hwang, *Applications and challenges of federated learning paradigm in the big data era with special emphasis on COVID-19*, Big Data and Cognitive Computing, **6**(4) (2022), 127. http://dx.doi.org/10.3390/bdcc6040127

[26] X. Mei, H. C. Lee, K. Y. Diao, M. Huang, B. Lin, C. Liu, Z. Xie, Y. Ma, P. M. Robson, M. Chung, A. Bernheim, V. Mani, C. Calcagno, K. Li, S. Li, H. Shan, J. Lv, T. Zhao, J. Xia, Q. Long, S. Steinberger, A. Jacobi, T. Deyer, M. Luksza, F. Liu, B. P. Little, Z. A. Fayad, Y. Yang, *Artificial intelligence–enabled rapid diagnosis of patients with COVID-19*, Nature Medicine, (2020). https://doi.org/10.1038/s41591-020-0931-3

[27] K. Michalak, H. Kwasnicka, *Correlation based feature selection method*, International Journal of Bio-Inspired Computation, **2**(5) (2010), 319-332. http://dx.doi.org/10.1504/IJBIC.2010.036158

[28] J. Mozaffari, A. Amirkhani, S. B. Shokouhi, *A survey on deep learning models for detection of COVID-19*, Neural Computing and Applications, (2023), 1-29. https://doi.org/10.1007/s00521-023-08683-x

[29] M. T. Nafees, I. M. Rizwan, M. M. I. Khan, M. Farhan, *A novel convolutional neural network for COVID-19 detection and classification using chest X-ray images*, medRxiv, (2021). https://doi.org/10.1101/2021.08.11.21261946

[30] A. Nandal, M. Blagojevic, D. Milosevic, A. Dhaka, L. N. Mishra, *Fuzzy enhancement and deep hash layer based neural network to detect Covid-19*, Journal of Intelligent and Fuzzy Systems, **41** (2021), 1341-1351. http://dx.doi.org/10.3233/JIFS-210222

[31] A. Narin, C. Kaya, Z. Pamuk, *Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks*, Pattern Analysis and Applications, **24**(3) (2021), 1207-1220. https://doi.org/10.1007/s10044-021-00984-y

[32] J. Ning, *Neural network-based pattern recognition in the framework of edge computing*, Romanian Journal of Information Science and Technology (ROMJIST), **27**(1) (2024), 106-119. http://dx.doi.org/10.59277/ROMJIST.2024.1.08

[33] S. Opricovic, *Multicriteria optimization of civil engineering systems*, PhD Thesis, Faculty of Civil Engineering, Belgrade, 1998.

[34] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, U. R. Acharya, *Automated detection of COVID-19 cases using deep neural networks with X-ray images*, Computers in Biology and Medicine, **121** (2020), 103792. https://doi.org/10.1016/j.compbiomed.2020.103792

[35] T. Rahman, M. E. H. Chowdhury, A. Khandakar, K. R. Islam, K. F. Islam, Z. B. Mahbub, M. A. Kadir, S. Kashem, *Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray*, Applied Sciences, **10**(9) (2020), 3233. http://dx.doi.org/10.3390/app10093233

[36] P. P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, A. Y. Ng, *CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning*, in arXiv:1711.05225, (2017). http://dx.doi.org/10.48550/arXiv.1711.05225

[37] C. H. Shin, J. J. Lee, C. Y. Jung, *An enhanced algorithm for an optimal high-frequency emphasis filter based on fuzzy logic for chest X-ray images*, Journal of Information and Communication Convergence Engineering, **13**(4) (2015), 264-269. http://dx.doi.org/10.6109/jicce.2015.13.4.264

[38] K. Simonyan, A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, in arXiv preprint arXiv:1409.1556, (2014). https://doi.org/10.48550/arXiv.1409.1556

[39] W. P. Sousa, C. C. P. Cruz, R. S. Lanzillotti, *Fuzzy divergence for lung radiography image enhancement*, Trends in Computational and Applied Mathematics, **24**(4) (2023), 699-716. http://dx.doi.org/10.5540/tcam.2023.024.04.00699

[40] H. Zeng, Y. M. Cheung, *Feature selection and kernel learning for local learning-based clustering*, IEEE Transactions on Pattern Analysis and Machine Intelligence, **33**(8) (2011), 1532-1547. http://dx.doi.org/10.1109/TPAMI.2010.215