IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 1, February 2026, pp. 592~603
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i1.pp592-603

Pneumonia classification from chest X-rays using significant feature selection and machine learning

Yugandhar Chodagam, Manjunatha Hiremath
Department of Computer Science, Central Campus, Christ University, Bengaluru, India

Article Info

Article history:

Received Aug 22, 2025
Revised Dec 27, 2025
Accepted Jan 10, 2026

Keywords:

Feature selection
Machine learning
Pneumonia
Radiomics
Wavelet features
X-ray

ABSTRACT

The chest X-ray images of normal lungs differ only subtly from those of lungs with pneumonia, making image-based diagnosis highly challenging. To address this issue, we developed a machine learning (ML)-based, lightweight, end-to-end Python package that processes chest X-ray images, implements robust feature selection methods, and classifies the images using various algorithms. While many studies have focused on improving classification accuracy using newer methods, few have addressed the interpretability of the extracted features or the growing computational demands of complex models. We used four publicly available datasets and extracted first-order, textural, and transform-based radiomic features to test our package. Features were selected using the Shapley additive explanations (SHAP) combined with recursive feature elimination (RFE) and stability selection algorithms. Our final solution contains a method that extracts a finite set of features identified by stability selection and feeds them as inputs into classical ML algorithms. Our model achieved 98% accuracy on the primary dataset, and 97%±1, 96%±2, and 94%±2% accuracy on the other three datasets. Our approach is fast, self-contained, and requires only an ideal set of features, making it suitable for resource-constrained clinical environments.

Corresponding Author:

Yugandhar Chodagam
Department of Computer Science, Central Campus, Christ University
Bengaluru, India
Email: yugandhar.ch@res.christuniversity.in

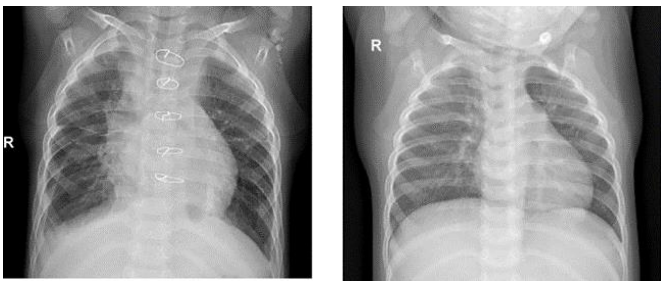
1. INTRODUCTION

Pneumonia remains one of the most serious health threats worldwide. According to a report published by the Center for Disease Control and prevention in 2023, pneumonia claims more than 41,000 lives every year in the United States, and this burden could be heavier in developing countries with poorer medical facilities. Distinguishing a healthy lung from one with pneumonia on a chest X-ray is often more difficult than it appears, even for a trained practitioner. Figure 1 shows representative chest X-ray images used in this study: Figure 1(a) illustrates a normal case, while Figure 1(b) illustrates pneumonia. Visual clues like patchy opacities or faint consolidations can be subtle and easy to miss, especially for less experienced radiologists. Patients who visit hospitals with limited resources face many hurdles when enough experienced radiologists are unavailable. A radiologist should have enough expertise in reading chest X-rays and identifying a pathology with an acceptable level of confidence. This dependency on experienced radiologists can be mitigated through automated image diagnosis or assisted diagnosis. Deep learning (DL) systems have already demonstrated impressive accuracy in many medical imaging tasks. However, they often rely on massive, well-annotated datasets and powerful hardware [1], [2]. Another promising solution is to combine machine learning (ML) algorithms with radiomics methods to accurately classify X-ray images by extracting

measurable and understandable features, such as texture, shape, and transform. Some powerful ML algorithms can learn from these informative sets of descriptors and achieve reliable performance with less computational demands [2].

This research aims to combine the strengths of radiomics and ML to pinpoint the most discriminative features using advanced feature selection methods. We propose a practical, high-performing packaged solution for pneumonia classification with a systematic approach and extensive evaluation across multiple datasets. Vinod *et al.* [3] report that their deep Covix-Net convolutional neural networks (CNNs) obtains patterns directly from the image pixels, and a separate pipeline first extracts texture and wavelet features and then lets a random forest algorithm learn from those descriptors. CNNs learn patterns from raw pixels, whereas radiomics methods extract texture and transform features, which can be fed into ML algorithms. Both these approaches have their strengths and limitations. Anderson *et al.* [4] designed a DL–based system and trained it on 490,000 chest X-rays, achieving an area under the curve (AUC) of 0.976 for pathology detection and improved accuracy on unseen held-out datasets. Verma *et al.* [1] used transform features such as wavelets to extract features such as entropy and energy, achieving an accuracy of 96.5% using a support vector machine algorithm, Khattab *et al*. [5] proposed focal loss–tuned Inception models, which demonstrated an overall accuracy of 97.67% with recall rates above 96%. Rabbah *et al.* [6] added three dense layers over the Inception-v3 extractor with approximately 22.9 million weights and reported an accuracy of 97.23% for binary pneumonia detection. Employing heterogeneous ensembles, efficient-VGG16 neural network achieved 99.46% accuracy on the covid-xray-5k benchmark [7], while Jaghdam *et al.* [8] achieved an even higher accuracy of 99.59% by hyper-optimizing an efficient DenseNet backbone. Surendra *et al.* [9] CXNet model demonstrated an accuracy of 98%, with a 96% recall rate for detecting coronavirus disease 2019 cases in a multiclass classification.

However, a non–DL approach can be equally effective in medical image classification. Özcan [10] lightGBM algorithm, trained on just 97 multiscale radiomic features, still delivered an accuracy, sensitivity, and specificity of 97.5, 97.5, and 98.75%, respectively. Guan *et al.* [11] proposed the use of a compact hash layer for extracting embeddings from the DenseNet-121 network, achieving an average precision of 0.84 for retrieving pneumonia cases. Pal *et al.* [12] designed a hybrid CNN-transformer model that yielded an accuracy of 95.14% when tested on the same three-class benchmark. These results show that while modern CNNs or CNN-vision transformer ensembles often achieve accuracies of 98-99%, features identified through stability selection and other carefully chosen radiomic features can match that performance with far less computational overhead.



(a) (b)

Figure 1. Representative chest X-ray samples of (a) normal and (b) pneumonia

2. METHOD

Researchers worldwide have expended considerable efforts for training computers to detect pneumonia and other thoracic diseases directly from X-rays using DL and ML methods. While the progress has been encouraging, the following issues have persisted: i) models trained on one dataset do not always perform well on another; ii) thousands of overlapping features can mask the truly useful signals; and iii) heavyweight architectures can be too slow or costly for everyday clinical use. To address these issues, we developed a self-contained package that can facilitate the extraction of significant features, enabling the classification of chest X-rays as normal or indicative of pneumonia with ideal accuracy levels. The designed package should ideally demonstrate stable and optimal performance across multiple datasets, even in settings with limited computational resources. The main advantages of this approach are that it: i) expedites image-based diagnosis, ii) relies on general computers that are easily available, and iii) assists beginner radiologists/annotators while reducing the burden on expert radiologists.

First, radiomic descriptors were initially extracted using the PyRadiomics package [13], which follows the image biomarker standardization initiative. Subsequently, the most informative features were identified and retained by applying two rigorous feature selection techniques namely, Shapley additive explanation (SHAP) combined with recursive feature elimination (RFE) and stability selection. A set of images was read into memory in the Google Colab environment along with the PyRadiomics package. Before the feature extraction, all the images were converted into nearly raw raster data (nrrd) files. Approximately 955 features were extracted, namely, original statistical features from raw images (95); transform features such as wavelet-low-pass, high-pass (LH) (86), wavelet- high-pass, low-pass (HL) (86), wavelet-high-pass, high-pass (HH) (86), wavelet-low-pass (LL) (86), exponential (86), gradient (86), logarithm (86), square (86), square root (86), and lbp-2d (86). These extracted features were fed into algorithms such as XGBoost, gradient boosting, and random forest to select the features of high importance across two datasets. A flowchart depicting the classification of images using SHAP and RFE is illustrated in Figure 2.

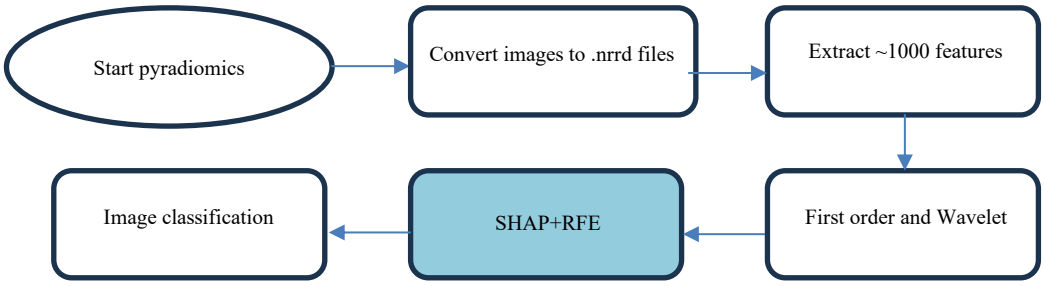


Figure 2. Flowchart of image classification using the SHAP and RFE methods

The four datasets used in this study were obtained from the following sources: Guangzhou Women and Children's Medical Center (GWCMC), Radiology AI Group (RAIG), National Institute of General Medical Sciences (NIGMS), and Jagannath University. SHAP combined with RFE exhibited excellent performance in identifying the features for each dataset separately. Although the models achieved satisfactory classification accuracy, the extracted feature sets across all the datasets shared only 30% common features. SHAP explains a single prediction by splitting the difference between the model output for that instance, f(x), and the model's expected output, E(f(z)), over a reference dataset [14]. Although the SHAP and RFE techniques yielded good accuracy for each dataset (GWCMC 98%; RAIG 98%; NIGMS 97%; Jagannath 94%), the SHAP features varied across datasets. The number of common features was low, and the SHAP values of the features were tied to the dataset.

Conversely, the flowchart in Figure 3 and the algorithm in Figure 4 achieved optimal classification accuracy and identified feature sets with over 90% commonality across all datasets. Optimal accuracy was achieved because stability selection draws many bootstrap subsets, fits a sparse model on each, and keeps only those variables whose selection frequency exceeds a user-defined threshold, which we set as 20% [15]. Table 1 presents data on the percentage of common features identified across the four datasets using different extraction methods.

The methods used for feature extraction are described in the following subsections. To keep the document concise, all the features extracted using SHAP combined with RFE are not listed here. This is because they are generally a subset of the features extracted using the stability selection algorithm.

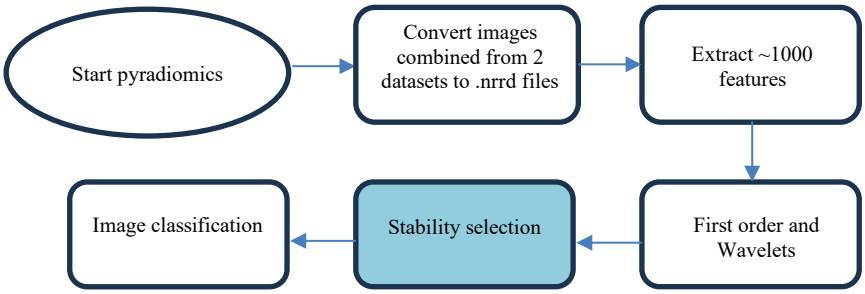


Figure 3. Image classification using stability selection

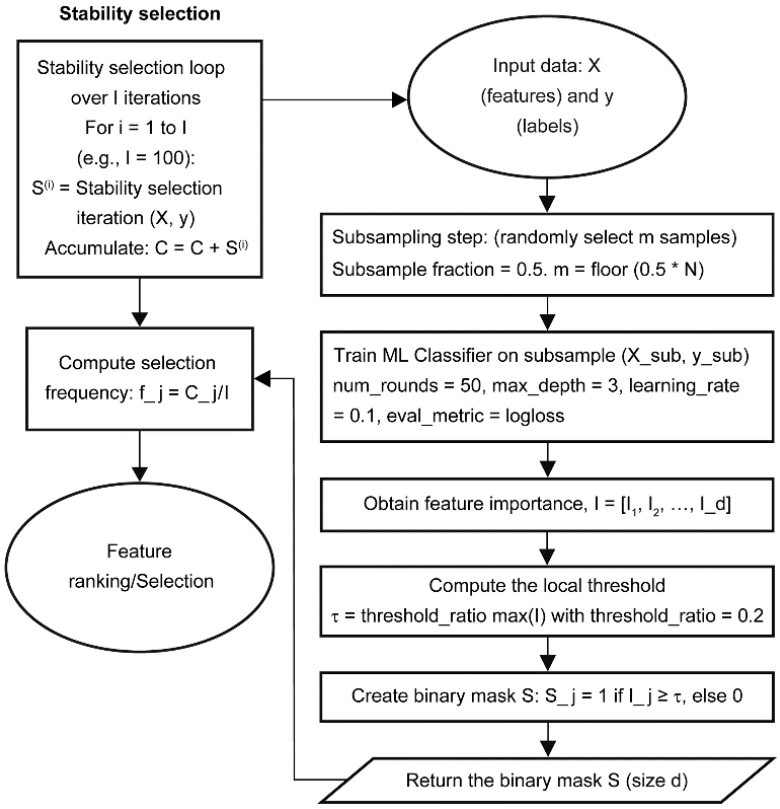


Figure 4. Stability selection algorithm

Table 1. Comparison of common feature percentages across four datasets using different methods

Method used	Feature match percentage across datasets (%)	Comments
SHAP+RFE	~30	Features common across three datasets went up to 30%
Stability selection using logistic regression	~70	Logistic regression in stability selection gave good results
Stability selection with XGBoost	>90	XGBoost produced the best results when the stability selection algorithm was used

2.1. Shapley additive explanations and recursive feature elimination algorithms

2.1.1. Shapley additive explanations

The Shapley function calculates a value for each feature extracted from each image. This value reflects the importance of that particular feature in the overall image classification. It is the average marginal contribution of a feature across all possible combinations.

$$\phi_i(f,x) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} \left(f_x(S \cup \{i\}) - f_x(S)\right) \quad (1)$$

Where f denotes the ML model; x denotes the instance for which we want to explain the prediction; S denotes a feature subset; $f_x(S)$ denotes the prediction for instance x using only the S features; and ϕ_i denotes the SHAP value for feature I; $\left(f_x(S \cup \{i\}) - f_x(S)\right)$ represents the marginal contribution of feature i when it is added to the subset S; and $|S|!\,(|F| - |S| - 1)!/|F|!$ denotes the weighting factor for the number of ways subset S can be formed, ensuring average contribution over all possible subsets.

2.1.2. Recursive feature elimination

RFE is a feature selection algorithm that aims to identify the most important features for a predictive model by removing the least important features recursively. Train the model as follows: train a model using all the dataset features.

$$M^{(t)} = Train\left(X^{(t)}, y\right) \quad (2)$$

Where $X^{(t)}$ denotes the feature matrix at iteration t, and $M^{(t)}$ denotes the trained model at iteration t.
Rank features: compute the importance of each feature in the trained model $M^{(t)}$. Let $\omega^{(t)} \epsilon \mathbb{R}^p$ be the vector of feature importance or coefficients.

$$\omega^{(t)} = FeatureImportance\left(M^{(t)}\right) \quad (3)$$

Eliminating the least important feature: feature with the smallest importance score is identified and removed from the feature set.

$$f^{(t)} = arg\,min\left(\omega_j^{(t)}\right) \quad (4)$$

2.2. Stability selection

The process of pneumonia classification using the stability selection algorithm is depicted in Figure 3. The workflow is almost identical to that associated with the SHAP+RFE method except it involves two datasets—GWCMC and a Kaggle dataset—and the algorithm to select the significant features is changed, denoted in blue color. A detailed flowchart describing the stability selection algorithm is illustrated in Figure 4. The algorithm works by repeatedly training an ML classifier on random data subsets. We used the logistic regression and XGBoost algorithms and tracked how often each feature was identified as important. This helps identify features that consistently contribute to classification across different samples. The features extracted using the XGBoost algorithm were found to facilitate the classification of images across all four datasets as shown in Table 2. This method helps select stable features across many iterations, increasing the robustness of the model while reducing overfitting. The algorithm presented in Figure 4 is explained as Algorithm 1.

Table 2. List of features identified through stability selection

No.	Feature name	No.	Feature name
1	wavelet-LH_glcm_ClusterTendency	13	wavelet-HH_glcm_Idn
2	wavelet-HH_glrlm_RunEntropy	14	wavelet-HH_glcm_Idmn
3	wavelet-LH_glrlm_ShortRunEmphasis	15	wavelet-HH_glcm_Idm
4	wavelet-LH_glrlm_RunEntropy	16	wavelet-HH_glcm_Imc1
5	wavelet-LH_glcm_SumEntropy	17	wavelet-HH_glcm_Imc2
6	wavelet-HH_glrlm_RunLengthNonUniformityNormalized	18	wavelet-LH_glcm_Id
7	wavelet-LH_glrlm_RunLengthNonUniformityNormalized	19	wavelet-LH_glcm_Idn
8	wavelet-LH_std	20	wavelet-LH_glcm_Idmn
9	wavelet-HH_energy	21	wavelet-LH_glcm_Idm
10	wavelet-HH_std	22	wavelet-LH_glcm_Imc1
11	wavelet-HH_mean	23	wavelet-LH_glcm_Imc2
12	wavelet-HH_glcm_Id		

Algorithm 1. Stability selection

i) Input: start with the merged datasets X (features) and y (labels).
ii) Subsampling: a random half of the dataset (set to 50%) is selected for each iteration.
iii) Model training: an ML model (XGBoost) is trained on the subsample with fixed parameters (e.g., 50 rounds, depth 3).
iv) Feature importance: after training, the model features are collected.
v) Thresholding: a local threshold is computed (set at 20% of the maximum importance in this study).
vi) Binary mask: a binary vector is created that evaluates a selected feature as 1 if it exceeds the threshold and 0 otherwise.
vii) Repeat: this process is repeated for multiple iterations (e.g., 100 times), and all binary vectors are accumulated.
viii) Final selection: the selection frequency for each feature is computed, and those that consistently appear are ranked and selected.

Table 2 presents a list of 23 discriminative features (out of the top 30) identified using the stability selection algorithm. The first seven features are indicators of the disorder or spread in the wavelet sub-bands. The next four features are related to the identification of strong edges in the infected field. Recognizing a

pattern, we looked for other descriptors that follow the same mathematical principles but were not included in the original list of 955 features. This was necessary because the accuracy of image classification in the Jagannath dataset was just around 91%.

Therefore, we systematically searched for additional features to improve the classification accuracy. We found that Gradient_entropy_mean is similar to other entropy measures, except that it is calculated over the Sobel gradient magnitude histogram. In addition, Sobel_edge_mean captures the average gradient magnitude, whereas Frangi_mean emphasizes elongated ridge-like structures. Incorporating Gradient_entropy_mean and Frangi_mean into our model increased the accuracy of image classification in the Jagannath dataset to 94%. Except for Gradient_entropy_mean, Sobel_edge_mean, and Frangi_mean, all the selected features are calculated after performing wavelet transformation. The wavelet features are specifically from the LH and HH sub-bands and primarily from two types of textures: gray level co-occurrence matrix and gray level run length matrix.

Using t-distributed stochastic neighbor embedding, we projected the top wavelet features from the GWCMC dataset into a two-dimensional space as presented in Figure 5. The blue and red points represent chest X-rays belonging to the normal and pneumonia categories, respectively. We can clearly observe that the selected features robustly capture the discriminative structure.

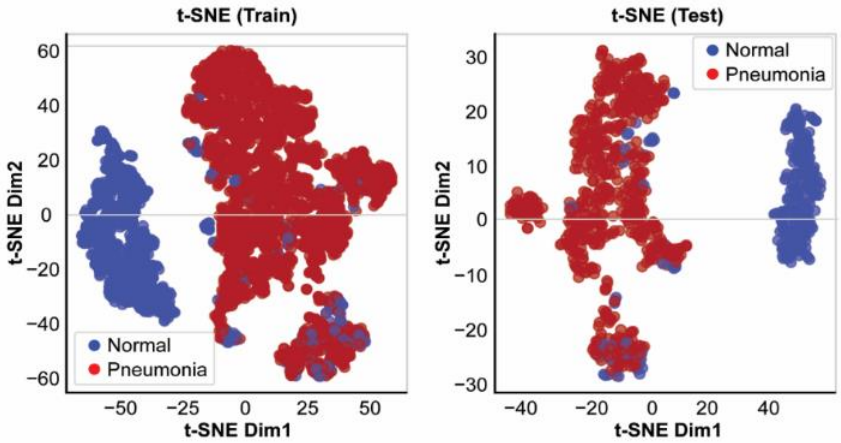


Figure 5. T-distributed stochastic neighbor embedding projection of the top wavelet features

2.3. Feature understanding

Anomalies or lesions are typically localized in medical images, and technicians, radiologists, and physicians focus on those regions to identify pathological biomarkers. Several techniques are employed in machine vision–based diagnostics, such as fast fourier transforms, Gabor filters, and wavelet transforms. Wavelet transforms are the most widely adopted because they capture both frequency and spatial information, whereas fast fourier transforms capture only the global frequency content by decomposing signals into sine and cosine components [16], [17].

First-order statistics extracted from the raw pixels of an image do not provide the detailed information required for image classification because they lack spatial and frequency contexts. Raw pixels are flat distributions of intensities and first-order statistics on such a distribution ignore edges, textures, and fine-scale patterns, features that are critical for medical image-based diagnosis. Isolating the fine edges, textures, or subtle variations in pixel intensity is key to detecting any abnormalities or pathologies in an image. Wavelets decompose the image into a set of sub-bands using wavelets or wave-like oscillations. Various frequencies and orientations are captured by each sub-band, and some of these details may capture the observed signal [18]. The image is first decomposed along rows and then along columns, producing four sub-bands: LL, which captures the coarse structure of the image; LH, which captures the vertical edges; HL, which captures the horizontal edges; and HH, which captures the diagonal edges and noise. The low-pass filters remove high-frequency details such as edges and noise, whereas the high-pass filters emphasize edges, boundaries, and fine details by removing the low-frequency, smooth background.

2.4. Feature explanation

Building on the observations mentioned above, we created a Python package to extract only the 26 aforementioned features from the images and classify the images using the XGBoost algorithm

(Figure 4). Other algorithms such as random forest were also tested, but XGBoost performed slightly better. This mirrors the findings in lung disease imaging studies where XGBoost achieved higher accuracy and better overall performance metrics than random forest [19]. Pneumowave package extracts the 26 features after minimal preprocessing of images and also does the classification of images and the flowchart is depicted in Figure 6.

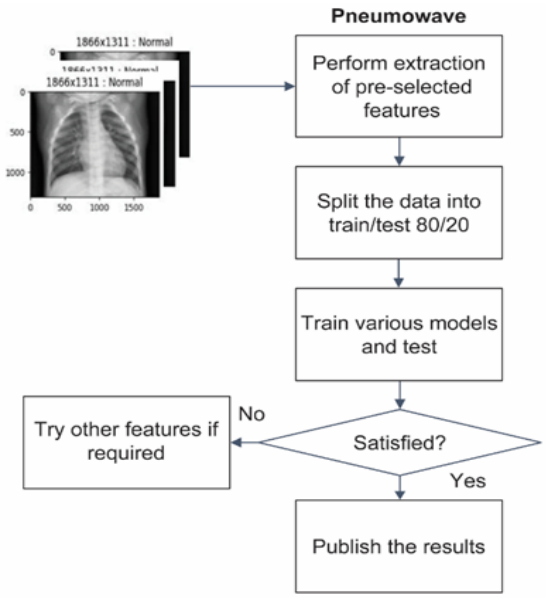


Figure 6. Pneumowave package used to extract 26 features and perform image classification

Some of the significant features are described as follows:

i) wavelet-LH_glcm_ClusterTendency: cluster tendency quantifies the extent to which a pair of pixels with similar intensities group together in a small region at a specific distance and orientation.

$$Cluster\ tendency = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \left(i + j - \mu_x - \mu_y\right)^2 P(i,j) \quad (5)$$

Where N_g denotes the mean gray-level of the sub-band, $P(i,j)$ denotes normalized gray-level co-occurrence matrix value, and μ_x, μ_y denotes mean gray-level of the reference pixels, neighbor pixels respectively.

ii) wavelet-HH_glrlm_RunEntropy: entropy is a measure of the unpredictability or randomness of a system. Run entropy describes the distribution of runs for sequences of consecutive pixels with the same intensity. This is useful for detecting regions with distortion, which might indicate a pathology. The sensitivity of wavelet-LH_glrlm_RunEntropy shifts toward horizontal structures.

$$Run\ entropy = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i,j)\ log\ P\ (i,j) \quad (6)$$

Where N_g denotes number of gray levels after you quantize the image, N_r denotes maximum run length and $P(i,j)$ denotes the probability of encountering a run of length j at gray-level i.

iii) wavelet-LH_glrlm_ShortRunEmphasis (SRE): SRE identifies fine-grained texture patterns, such as microtextural changes and subtle edges, by assigning higher weights to shorter runs.

$$Short\ run\ emphasis = \frac{1}{N_r} \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i,j)}{j^2} \quad (7)$$

Where N_g denotes number of gray levels after you quantize the image, N_r denotes maximum run length, and j^2 denotes squared run length in the denominator, giving short runs a larger weight than long runs.

iv) wavelet-LH_glcm_SumEntropy: sum entropy is similar to run entropy except that it is the sum of gray-level values of neighboring pixels, which captures the texture complexity in the image.

$$Sum\ entropy = -\sum_{k=2}^{2N_g} p_{x+y}(k)\ log\ p_{x+y}(k) \quad (8)$$

Where N_g denotes number of gray levels after you quantize the image, and $p_{x+y}(k)$ denotes sum distribution of the GLCM.

v) wavelet-HH_glrlm_RunLengthNonUniformityNormalized (RLNUN): run length non-uniformity normalized measures the variation in the lengths of pixel runs in an image. It tells us whether the runs (i.e., sequences of pixels with the same intensity) are mostly of the same length or spread across different lengths. The wavelet-LH_glrlm_RunLengthNonUniformityNormalized feature captures the same measurement for the LH sub-band.

$$RLNUN = \frac{1}{N_r}\sum_{j=1}^{N_r}\left(\sum_{i=1}^{N_g} P(i,j)\right)^2 \quad (9)$$

Where N_g denotes number of gray levels after you quantize the image, N_r denotes maximum run length, and $P(i,j)$ denotes the probability of encountering a run of length j at gray-level i.

vi) wavelet-LH_std (STD): STD quantifies how widely pixel values deviate from the mean, making it sensitive to local variations in brightness and texture.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \quad (10)$$

Where σ denotes standard deviation, x_i denotes the ith observation in the sample, and μ denotes average.

vii) wavelet-HH_energy: energy serves as an indicator of the presence and magnitude of detailed structures, such as fine edges, speckle noise, or textural patterns.

$$Energy = \sum_{i=1}^{N} x_i^2 \quad (11)$$

Where N denotes the total number of samples, and x_i denotes the ith observation in the sample.

viii) wavelet-HH_mean: mean provides information about the overall brightness or signal level in the diagonal texture regions.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (12)$$

Where μ deontes the average and x_i denotes the ith observation in the sample.

ix) gradient_entropy_mean (GEM): GEM captures the unpredictability of gradient changes in the image. Chest X-rays of lungs with pneumonia present disordered textures and uneven transitions, which are detected well by this feature.

$$GEM = {}^{1}/_{N} \cdot \sum_{j=1}^{N}\left(-\sum p_i^{(j)} \cdot log_2\left(p_i^{(j)}\right)\right) \quad (13)$$

Where N denotes the total number of neighborhoods over which entropy is averaged and $p_i^{(j)}$ denotes normalized histogram bin.

3. RESULTS AND DISCUSSION

The GWCMC dataset was our primary dataset for image classification as shown in Table 3. We named the package we designed as Pneumowave. Pneumowave included minimal preprocessing of data (images were resized to 224×224 and pixel intensities were normalized), extraction of 26 features (based on the features identified by the stability selection algorithm and more), model building (XGBoost with max_depth: 6, learning_rate: 0.05, n_estimators: 300), and testing. The package achieved an accuracy of 98% and a sensitivity of 99% on the GWCMC dataset. The top 11 of the 26 features listed in Table 2 were used in the model.

Researchers have widely used the GWCMC dataset for pneumonia detection. The deep CNN model built by Tang *et al.* [20] demonstrated an accuracy of 94.64%, sensitivity of 96.5%, and specificity of

92.86%. Wang *et al.* [21] established a DL pipeline that achieved an AUC in the range of 0.94–0.98 for pneumonia classification. Prakash *et al.* [22] proposed transfer learning with stacked CNNs, achieving 96.15% accuracy, 97.90% precision, and 95.90% recall. Gupta *et al.* [23] used neural architecture search and reported a good AUC of 0.976, but their approach required high computational effort. Shati *et al.* [24] combined wavelets, gray level co-occurrence matrix, and k-nearest neighbors to achieve a high AUC and an accuracy of nearly 99%; however, they used a very complex model. Singh *et al.* [25] applied vision transformers, achieving a high accuracy of 97.61% and an AUC of 0.96. Raghaw *et al*. [26] used dilated convolution and transformers, obtaining an accuracy of 92.87%, precision of 91.72%, and recall of approximately 91.59%. However, both these approaches were complex and computationally intensive. Our own DL model based on DenseNet-121 yielded an accuracy of 98.2% with minimal preprocessing, running for approximately 8 hours on a CPU-based system.

Table 3. Datasets used in this study

Dataset	Source	Image count	Compiled by
GWCMC	Guangzhou Women and Children's Medical Center, Guangzhou, China	5,856	Daniel *et al.*; https://data.mendeley.com/datasets/rscbjbr9sj/2 https://www.cell.com/cell/fulltext/S0092-8674(18)30154-5
RAIG	Italian Society of Medical and Interventional Radiology (SIRM)	2,998	Radiology AI Group, Qatar University and the University of Dhaka along with collaborators from Pakistan and Malaysia and some medical doctors; https://data.niaid.nih.gov/resources?id=figshare_14151854 https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database(Italian- SIRM)
NIGMS	NIGMS/NIH (National Institute of General Medical Sciences) (https://www.medrxiv.org/content/10.1101/2020.05.09.20096560v2)	240	Nature and Scientific Reports; https://github.com/abzargar/COVID-Classifier
Jagannath University	Various hospitals in Bangladesh	2,050	https://data.mendeley.com/datasets/m4s2jn3csb/1

When the top 11 features listed in Table 2 were used, Pneumowave achieved an accuracy of 98% on the GWCMC and RAIG datasets and 96% on the NIGMS dataset. The performance is illustrated in confusion matrices as shown in Figure 7. The accuracy of image classification in the NIGMS dataset improved to 98% when wavelet-HH_glcm_Idn and wavelet-LH_glrlm_graylevelNonUniformityNormalized were added as features. Similarly, the classification accuracy in the Jagannath dataset increased up to 94% when Gradient_entropy_mean and Frangi_mean were added to the top 11 features. We tried to incorporate few image preprocessing techniques, including denoising, contrast-limited adaptive histogram equalization (a technique used to enhance local image contrast), and Otsu thresholding. However, they resulted in reduced performance because these methods might inadvertently remove the relevant details or segments. In any case, the reduced accuracy observed in the Jagannath dataset can be partly attributed to the inferior image quality, an aspect examined in detail in the following section.

3.1. Quality of images

The accuracy of image classification depends heavily on the quality of images. Images from the GWCMC, RAIG, and NIGMS datasets had better quality than those from the Jagannath dataset. The mean height of images from the Jagannath dataset was 299 pixels, whereas that of images from the other three datasets ranged from 820 to 1,472 pixels. Similarly, the width of images from the Jagannath dataset was 299 pixels, whereas that of images from the other three datasets ranged from 1,039 to 1,686 pixels.

The mean size of images from the Jagannath dataset was 10 kB, whereas that of images from the other three datasets ranged from 81 to 538 kB. The Tenengrad sharpness of images from the GWCMC, RAIG, and NIGMS datasets was 0.77, 0.96, and 0.84, respectively, indicating a good balance between images in the normal and pneumonia categories. The Tenengrad sharpness of images from the Jagannath dataset was as low as 0.29, indicating that normal images from this dataset were much less sharp than images of lungs with pneumonia. This imbalance could be responsible for the poor classification performance, as also noted by Pollatou and Ferrante [27].

3.2. Pneumonia classification performance of pneumowave package

The features identified using the stability selection algorithm listed in Table 2 were extracted for images in three other datasets using the developed Pneumowave package. Table 4 presents the data on the

count of images, accuracy, sensitivity, specificity, and AUC in the final datasets. In addition to this, performance of our model is either comparable to or, in some cases, even better than the most advanced DL architectures mentioned in the results section for pneumonia classification on the GWCMC dataset.

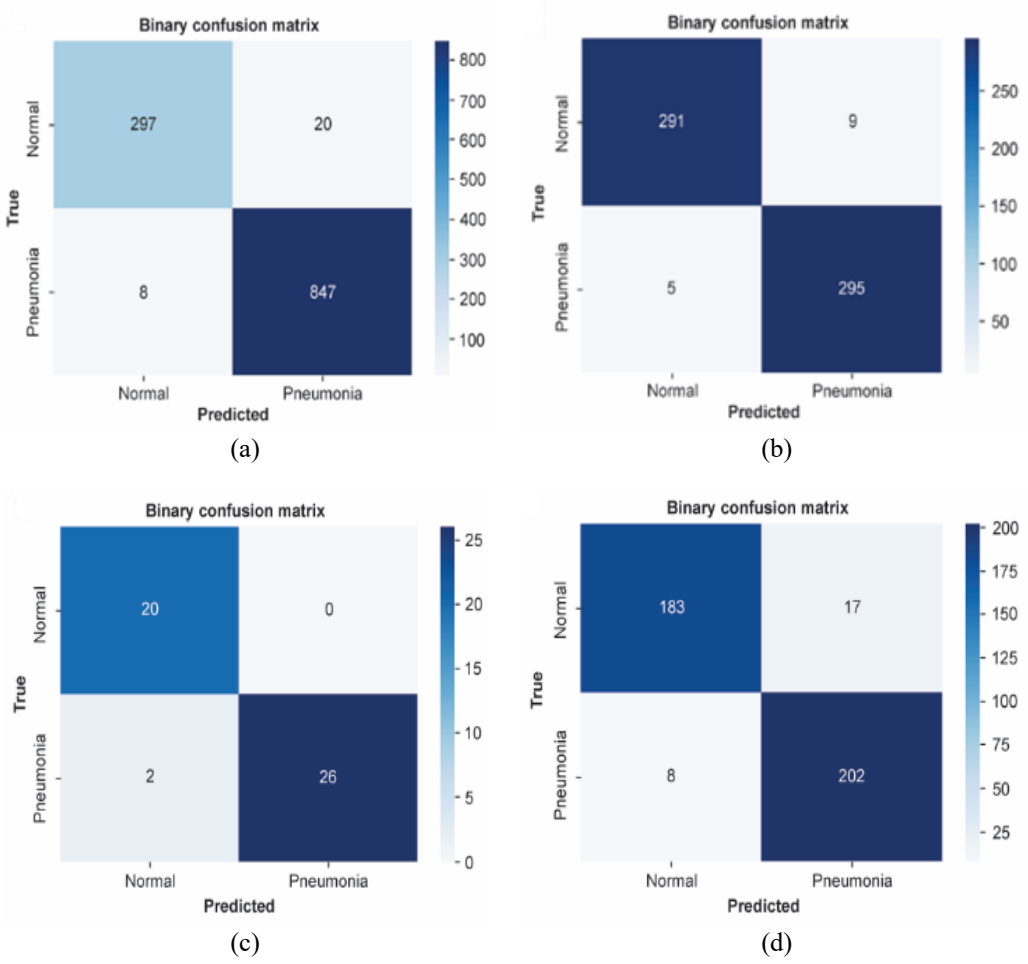


Figure 7. Confusion matrices of (a) GWCMC, (b) RAIG, (c) NIGMS, and (d) Jagannath

Table 4. Various datasets and classification results

Dataset	Train (n)	Test (n)	Total	Accuracy	Sensitivity	Specificity	AUC
GWCMC	4684	1172	5856	0.98	0.99	0.94	0.99
RAIG	2398	600	2998	0.98	0.98	0.97	0.99
NIGMS	192	48	240	0.96	0.93	1	0.98
Jagannath	1640	410	2050	0.94	0.96	0.92	0.98

4. CONCLUSION

This study demonstrates that the extraction and selection of a simple set of custom features combined with classical ML methods can deliver a classification accuracy comparable to that obtained using deep neural networks. This study extracted first-order statistics, texture, and wavelet measures from each X-ray image. Two selectors were tested, and stability selection paired with XGBoost proved to be the best feature extraction method. Stability selection identified 11 features that were sufficient across all four datasets and delivered accuracies of 98%, 98%, 96%, and 94% on the GWCMC, RAIG, NIGMS, and Jagannath datasets, respectively. Model training was completed in under a minute on standard hardware, and inference required no more than 13 variables. This approach is suitable for scenarios where resources are severely limited and heavy models are expensive to host. The results were similar to those of recent studies that combined classical texture cues with ML instead of a full deep network. This approach can be extended to multiclass classification that involves bacterial, viral, and coronavirus disease 2019 image classes or binary

classification of other pathologies. In conclusion, our study successfully demonstrated that careful feature selection, combined with an appropriate ML algorithm, enabled a fast and portable solution for pneumonia screening without relying on deep neural networks.

FUNDING INFORMATION

Authors state there is no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Yugandhar Chodagam	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓			
Manjunatha Hiremath	✓	✓			✓			✓		✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The authors confirm that the data supporting the findings of this study are available within the article

REFERENCES

[1] A. K. Verma, I. Vamsi, P. Saurabh, R. Sudha, S. G. R., and R. S., "Wavelet and deep learning-based detection of SARS-nCoV from thoracic X-ray images for rapid and efficient testing," *Expert Systems with Applications*, vol. 185, Dec. 2021, doi: 10.1016/j.eswa.2021.115650.

[2] C. O.- Toro, A. G.- Pedrero, M. L.- Saavedra, and C. G.- Martín, "Automatic detection of pneumonia in chest X-ray images using textural features," *Computers in Biology and Medicine*, vol. 145, Jun. 2022, doi: 10.1016/j.compbiomed.2022.105466.

[3] D. N. Vinod, B. R. Jeyavadhanam, A. M. Zungeru, and S. R. S. Prabaharan, "Fully automated unified prognosis of COVID-19 chest X-ray/CT scan images using deep Covix-Net model," *Computers in Biology and Medicine*, vol. 136, Sep. 2021, doi: 10.1016/j.compbiomed.2021.104729.

[4] P. G. Anderson *et al.*, "Deep learning improves physician accuracy in the comprehensive detection of abnormalities on chest X-rays," *Scientific Reports*, vol. 14, no. 1, Oct. 2024, doi: 10.1038/s41598-024-76608-2.

[5] R. Khattab, I. R. Abdelmaksoud, and S. Abdelrazek, "Automated detection of COVID-19 and pneumonia diseases using data mining and transfer learning algorithms with focal loss from chest X-ray images," *Applied Soft Computing*, vol. 162, Sep. 2024, doi: 10.1016/j.asoc.2024.111806.

[6] J. Rabbah, M. Ridouani, and L. Hassouni, "Improving pneumonia diagnosis with high-accuracy CNN-based chest X-ray image classification and integrated gradient," *Biomedical Signal Processing and Control*, vol. 101, Mar. 2025, doi: 10.1016/j.bspc.2024.107239.

[7] S. Kumar and H. Kumar, "Efficient-VGG16: a novel ensemble method for the classification of COVID-19 X-ray images in contrast to machine and transfer learning," *Procedia Computer Science*, vol. 235, pp. 1289–1299, 2024, doi: 10.1016/j.procs.2024.04.122.

[8] I. H.- Jaghdam *et al.*, "Towards COVID-19 detection and classification using optimal efficient densenet model on chest X-ray images," *Alexandria Engineering Journal*, vol. 101, pp. 136–146, Aug. 2024, doi: 10.1016/j.aej.2024.05.073.

[9] Surendra, M. V. M. Kumar, S. L. S. Darshan, and B. S. Prashanth, "CXNet-a novel approach for covid-19 detection and classification using chest X-ray image," *Procedia Computer Science*, vol. 235, pp. 2486–2497, 2024, doi: 10.1016/j.procs.2024.04.234.

[10] H. Özcan, "Exploring best-performing radiomic features with combined multilevel discrete wavelet decompositions for multiclass COVID-19 classification using chest X-ray images," *Computers in Biology and Medicine*, vol. 193, Jul. 2025, doi: 10.1016/j.compbiomed.2025.110392.

[11] A. Guan, L. Liu, X. Fu, and L. Liu, "Precision medical image hash retrieval by interpretability and feature fusion," *Computer Methods and Programs in Biomedicine*, vol. 222, Jul. 2022, doi: 10.1016/j.cmpb.2022.106945.

[12] M. Pal, S. Parija, and G. Panda, "An effective ensemble approach for classification of chest X-ray images having symptoms of COVID: a precautionary measure for the COVID-19 subvariants," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 8, Jun. 2024, doi: 10.1016/j.prime.2024.100547.

[13] J. J. M. V. Griethuysen *et al.*, "Computational radiomics system to decode the radiographic phenotype," *Cancer Research*, vol. 77, no. 21, pp. e104–e107, Nov. 2017, doi: 10.1158/0008-5472.CAN-17-0339.
[14] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, California, USA: Curran Associates Inc., 2017, pp. 4768–4777, doi: 10.5555/3295222.3295230.
[15] V. Hamer and P. Dupont, "An importance-weighted feature selection stability measure," *Journal of Machine Learning Research*, vol. 22, no. 116, pp. 1–57, 2021.
[16] C. Scapicchio, M. Gabelloni, A. Barucci, D. Cioni, L. Saba, and E. Neri, "A deep look into radiomics," *La radiologia medica*, vol. 126, no. 10, pp. 1296–1311, Oct. 2021, doi: 10.1007/s11547-021-01389-x.
[17] M. E. Mayerhoefer *et al.*, "Introduction to radiomics," *Journal of Nuclear Medicine*, vol. 61, no. 4, pp. 488–495, Apr. 2020, doi: 10.2967/jnumed.118.222893.
[18] H. Li and K. Xu, "Innovative adaptive edge detection for noisy images using wavelet and Gaussian method," *Scientific Reports*, vol. 15, no. 1, Feb. 2025, doi: 10.1038/s41598-025-86860-9.
[19] Z. Cao *et al.*, "Development, deployment, and feature interpretability of a three-class prediction model for pulmonary diseases," *Insights into Imaging*, vol. 16, no. 1, Jun. 2025, doi: 10.1186/s13244-025-02020-7.
[20] Y.-X. Tang *et al.*, "Automated abnormality classification of chest radiographs using deep convolutional neural networks," *npj Digital Medicine*, vol. 3, no. 1, May 2020, doi: 10.1038/s41746-020-0273-z.
[21] G. Wang *et al.*, "A deep-learning pipeline for the diagnosis and discrimination of viral, non-viral and covid-19 pneumonia from chest X-ray images," *Nature Biomedical Engineering*, vol. 5, no. 6, pp. 509–521, Apr. 2021, doi: 10.1038/s41551-021-00704-1.
[22] A. J. Prakash *et al.*, "Transfer learning approach for pediatric pneumonia diagnosis using channel attention deep CNN architectures," *Engineering Applications of Artificial Intelligence*, vol. 123, Aug. 2023, doi: 10.1016/j.engappai.2023.106416.
[23] A. Gupta, P. Sheth, and P. Xie, "Neural architecture search for pneumonia diagnosis from chest X-rays," *Scientific Reports*, vol. 12, no. 1, Jul. 2022, doi: 10.1038/s41598-022-15341-0.
[24] A. Shati, G. M. Hassan, and A. Datta, "A comprehensive fusion model for improved pneumonia prediction based on KNN-wavelet-GLCM and a residual network," *Intelligent Systems with Applications*, vol. 26, Jun. 2025, doi: 10.1016/j.iswa.2025.200492.
[25] S. Singh, M. Kumar, A. Kumar, B. K. Verma, K. Abhishek, and S. Selvarajan, "Efficient pneumonia detection using vision transformers on chest X-rays," *Scientific Reports*, vol. 14, no. 1, Jan. 2024, doi: 10.1038/s41598-024-52703-2.
[26] C. S. Raghaw, P. S. Bhore, M. Z. U. Rehman, and N. Kumar, "An explainable contrastive-based dilated convolutional network with transformer for pediatric pneumonia detection," *Applied Soft Computing*, vol. 167, Dec. 2024, doi: 10.1016/j.asoc.2024.112258.
[27] A. Pollatou and D. D. Ferrante, "Out-of-focus brain image detection in serial tissue sections," *Journal of Neuroscience Methods*, vol. 345, Nov. 2020, doi: 10.1016/j.jneumeth.2020.108852.

BIOGRAPHIES OF AUTHORS

Yugandhar Chodagam holds MBA in IT and GIS from University of Redlands, California, United States and is working as an assistant professor at Vignana Jyothi Institute of Management (Autonomous), Hyderabad, India. He is presently pursuing his Ph.D. at Christ University, Bengaluru. His research interests include medical image classification, radiomics, machine learning, and advanced feature selection methods for healthcare applications. He has over 25 years of experience in implementing IT projects and brings strong industry expertise to academics. He can be contacted at email: yugandhar.ch@res.christuniversity.in.

Manjunatha Hiremath is an assistant professor in Department of Computer Science at Christ University, Bengaluru. He has written over 35 journal articles in various international journals and his interest areas include computer vision, image and video processing, pattern recognition, biometrics. He is involved in several activities as a supervisor for many scholars who are pursuing their doctoral programmes and as a principal investigator for face dataset project along with many other academic duties. He can be contacted at email: manjunatha.hiremath@christuniversity.in.