Phi: Jurnal Pendidikan Fisika dan Terapan
Vol.11 No.2 (2025) 110-122
ISSN: 2460-4348, e-ISSN: 2549-7162
https://jurnal.ar-raniry.ac.id/index.php/jurnalphi/index

Article Type: *orginial research*

High-Accuracy Pneumonia Classification via Ensemble Learning on Chest X-ray Imagery

Rezky Rachmadany Rachman[1*], Syamsir Dewang[2], Sri Dewi Astuti[2], Eko Juarlin[2]
[1]Physics Department, Lambung Mangkurat University, Banjarbaru, Indonesia
[2]Physics Department, Hasanuddin University, Makassar, Indonesia

Correspondence: E-mail: rezkyrrachman@ulm.ac.id

ARTICLE INFO

Article History:
Submitted/Received: 29 April 2025
First Revised: 10 June 2025
Accepted: 14 June 2025
First Available Online: 01 July 2025
Publication Date: 01 July 2025

Keywords:
Pneumonia; Chest X-Ray; Image classification; HOG; LBP; Ensemble learning

ABSTRACT

Pneumonia continues to pose a substantial global health threat, necessitating rapid and precise diagnostic tools. The conventional manual assessment of Chest X-ray (CXR) images is time-intensive and susceptible to human error. This study introduces an automated machine learning approach that employs an ensemble learning strategy to achieve highly accurate pneumonia classification from CXR images. The comprehensive system operates through three primary phases: initial image pre-processing (involving grayscale conversion, resizing, and filtering for enhanced quality), robust feature extraction (utilizing the fusion of Histogram of Oriented Gradients (HOG) and Local Binary Pattern (LBP) descriptors), and sophisticated model training and classification. An ensemble model is trained, integrating the predictive power of Random Forest, Logistic Regression, and Extreme Gradient Boosting classifiers. Experimental validations, performed on a dedicated dataset comprising pneumonia and normal CXR images, unequivocally demonstrate that the proposed strategy achieves an impressive 97.50% overall classification accuracy, strongly supported by precision, recall, and F1-scores all at 97.50%. This superior performance, notably surpassing individual machine learning algorithms, underscores the profound efficacy of ensemble learning in delivering reliable and precise predictions for pneumonia diagnosis. Consequently, this automated methodology presents a valuable asset for medical professionals, aiding in the swift and accurate identification of pneumonia.

1. INTRODUCTION

Pneumonia, an acute respiratory infection impacting the lungs, continues to be a primary contributor to illness and mortality globally, particularly among vulnerable groups like children and the elderly. Swift and precise diagnosis of pneumonia is vital for effective patient management, enabling timely intervention and preventing severe complications. Chest X-ray (CXR) imaging stands as a fundamental diagnostic instrument due to its wide availability, cost-efficiency, and non-invasive nature.

However, the manual evaluation of CXR images by radiologists is a convoluted and laborious process demanding considerable expertise, and is susceptible to human errors, potentially leading to incorrect or delayed diagnoses. This issue is aggravated in areas lacking sufficient trained radiologists. Moreover, misdiagnosing pneumonia might result in improper treatment, as its symptoms can sometimes mimic those of other pulmonary conditions. This underscores an urgent requirement for accurate and rapid diagnostic methodologies.

Rapid advancements in artificial intelligence (AI) and machine learning (ML) have revealed promising avenues for automating and enhancing medical image analysis. ML techniques possess the capability to overcome the limitations of manual CXR analysis by offering objective, consistent, and prompt diagnostic assistance. Notably, ML approaches have been extensively employed in the healthcare sector for significantly accelerating and improving the precision of infection prediction (Alqaduh et al., 2021; Yadav et al., 2020; Batista et al., 2020). Researchers have actively explored diverse ML-driven methods for disease classification using medical imaging.

For instance, Rasheed et al. examined Logistic Regression (LR) and CNN for accurate automated diagnoses derived from X-ray images, achieving an overall accuracy of 95.2% with PCA (Rasheed et al. 2021). Saygili et al. introduced a technique to efficiently and accurately categorize respiratory conditions on Chest CT and X-ray images through image processing and ML, where their LBP+KNN model demonstrated detection accuracies ranging from 89.41% (Saygili et al., 2021). Mary and Raj sought to pinpoint optimal classification strategies for respiratory cases through various ML techniques, with SVM notably achieving an 85% accuracy (Mary and Raj 2021). Khuzani et al. developed an ML classifier that utilized a dimensionality reduction method and CXR images to precisely identify lung infections, showcasing good accuracy and sensitivity (Khuzani et al., 2021). Mijwil et al. presented an ML technique incorporating Random Forest, Logistic Regression, Naive Bayes, and SVM, which achieved a 91.8% accuracy with SVM on chest X-ray images for viral pneumonia and healthy cases (Mijwil et al., 2021). Alquran et al. utilized image processing techniques to extract textural features from radiological images, ultimately achieving a total accuracy of 93.1% with an Ensemble classifier (Alquran et al., 2021). Johri et al. proposed a novel approach for lung infection classification using a Random Forest algorithm on CXR images, reaching 92.4% accuracy (Johri et al., 2021). Furthermore, deep learning methods have gained significant traction, with studies such as those by Khan et al. (2020), Narin et al. and Rahman et al. (2020) employing various Convolutional Neural Network (CNN) architectures for COVID-19 detection from X-ray images (Narin et al., 2021; Rahman et al.,2020). Hybrid approaches, combining elements of deep learning or advanced processing, have also shown promise (Aslan and Sabanci, 2021; Karim and Rahman, 2022; Saraiva et al., 2021; Ulhaq et al., 2022). Additionally, recent work by Rezky R.R, et al. achieving 99.6% accuracy with K-NN for SARS-CoV-2 detection via GLCM-based feature extraction from CT-Scan images (Rezky R.R, et al., 2024). The study, although centered on SARS-CoV-2, underscores the significant potential of texture-based feature extraction and traditional ML classifiers for lung disease detection using CT-Scan images, a principle that also applies to pneumonia. Despite these extensive endeavors, certain conventional ML algorithms can still encounter misclassification challenges, especially with noisy datasets. This underscores the necessity for developing more robust and precise classification frameworks specifically designed for pneumonia detection.

The key contributions of this study are multifaceted. We propose a novel and effective ensemble-based machine learning approach designed specifically for the high-accuracy classification of pneumonia utilizing CXR images. This method employs an efficient fusion descriptor, combining Histogram of Oriented Gradients (HOG) and Local Binary Pattern (LBP) to extract significant and discriminative features from the images. Leveraging these prominent features, our proposed ensemble model accurately classifies pneumonia patients from normal cases. Furthermore, we rigorously validate the method's performance by comparing it against several individual state-of-the-art ML classifiers on the same CXR image dataset, thereby demonstrating its superiority across key metrics such as accuracy, precision, recall, and F1-score.

Building upon these advancements, this research presents a groundbreaking automated ML-driven system explicitly developed for high-accuracy pneumonia classification from CXR images. Its paramount novelty lies in the synergistic integration of a uniquely formulated feature fusion strategy with a robust ensemble learning paradigm. Specifically, our methodology employs a distinctive combination of Histogram of Oriented Gradients (HOG) and Local Binary Pattern (LBP) descriptors—an efficient fusion technique known for capturing both intricate shape and rich textural information—to extract highly discriminative features. These advanced features then feed into a meticulously structured ensemble of diverse individual classifiers (namely, Random Forest, Logistic Regression, and Extreme Gradient Boosting). This integrated approach is engineered to leverage the complementary strengths of different feature representations and algorithmic biases, thereby yielding significantly more dependable and reliable predictions than conventional single models, directly addressing critical challenges in current pneumonia diagnosis. This study holds particular urgency and importance due to ongoing global health challenges and the demand for efficient diagnostic tools to support healthcare infrastructures, especially in underserved regions where traditional diagnostic capacities may be strained. Automating a vital diagnostic step through this research holds the potential to ease the workload on healthcare practitioners and enhance patient recovery.

2. METHODS

2.1 Research Approach and Framework

Adopting a quantitative, applied research paradigm, this study focuses on distinguishing pneumonia from non-pneumonia cases in CXR images through numerical and statistical analysis. Its primary aim is to establish a viable alternative diagnostic system, leveraging integrated digital image processing and machine learning to generate precise diagnostic insights for expedited clinical intervention.

2.2 Data Collection and Preparation

The dataset central to this research consists of Chest X-ray (CXR) images. To mitigate bias and ensure representativeness, a meticulously balanced and randomly sampled collection was curated. Specifically, the dataset comprises a total of 400 CXR images**:** 200 samples depicting pneumonia**, and** 200 samples representing normal chest conditions. These images were obtained from the publicly accessible Pneumonia database on Kaggle. Initial quality control procedures were applied to discard illegible or suboptimal images. Furthermore, independent evaluations by two experienced physicians were conducted, with diagnoses corroborated by clinical history, symptom assessment, and laboratory findings. The complete dataset was then partitioned for training and testing, with 80% (320 images) allocated for model development and 20% (80 images) reserved for performance validation.

2.3 Research Method and Design

The intended framework for categorizing pneumonia is systematically divided into three main stages: preprocessing, feature extraction, and training and classification.

2.3.1 Pre-processing

During this stage, CXR images undergo a conversion to grayscale, primarily to streamline computational demands and highlight relevant image attributes. Subsequently, all images are uniformly scaled to a standardized input resolution of 100×100 pixels, aiming to normalize dimensions and further enhance computational efficiency. To boost image clarity and diminish noise artifacts, median and mean filters are systematically applied. This rigorous pre-processing regimen is crucial for ensuring the stability and effective convergence of the algorithms employed in later learning stages.

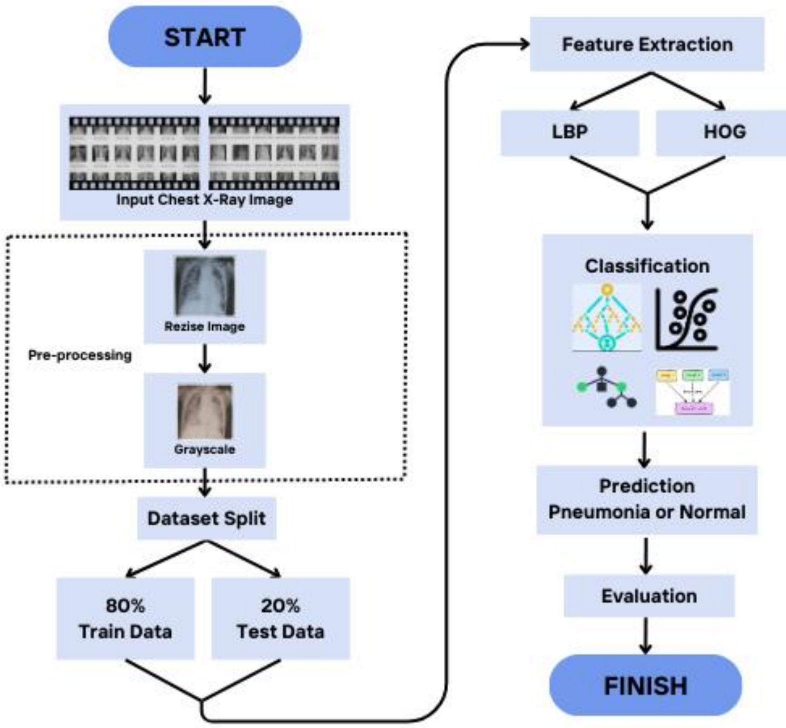


Figure 1. Flowchart of Pneumonia Classification

2.3.2 Feature Extraction

The process of feature extraction is tasked with identifying and isolating relevant patterns from the input data. In this investigation, two effective feature descriptors, the Histogram of Oriented Gradients (HOG) and Local Binary Pattern (LBP), are employed to derive distinct features from the pre-processed CXR images. These descriptors are chosen for their capability to capture complementary characteristics of images.

Histogram of Oriented Gradients (HOG)

HOG functions as a feature extractor that generalizes object representations by examining the distribution of intensity gradients or edge directions. This technique excels at extracting shape-related attributes from visual data. Gradients are computed within discrete blocks across an image, reflecting variations in pixel intensity magnitude and orientation. The magnitude of each pixel's gradient is calculated using the formula:

$$G = \sqrt{G_x^2 + G_y^2} \tag{1}$$

while its direction is determined by

$$\theta = arct\ an \frac{g_x}{g_y} \tag{2}$$

Here, G_x indicates the horizontal component of pixel intensity variations, and G_Y represents the vertical component. HOG descriptors are typically derived from 8×8 pixel blocks, with each block's directional angles and values summarized in a 9-bin histogram. To lessen the impact of varying lighting conditions, histograms undergo normalization across a 16×16 block size, which involves collectively normalizing four 8×8 blocks. This strategy helps minimize accuracy degradation due to illumination changes.

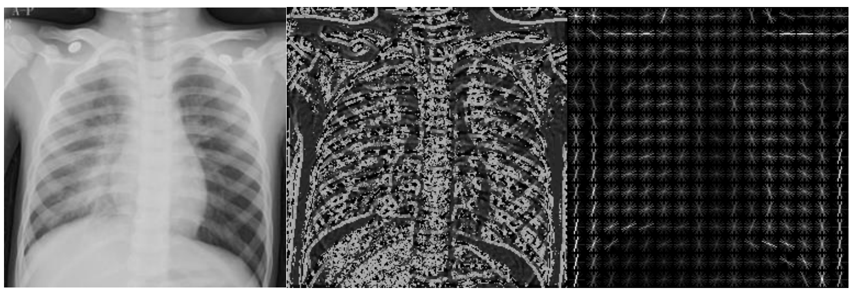
Local Binary Pattern (LBP)

LBP serves as a texture descriptor, operating on the differences between neighboring pixels and a central pixel. It generates a binary code for a pixel's value by applying a threshold based on its central pixel to its surrounding pixels. The mathematical representation for LBP is provided by Equation (3):

$$LBP(x_c, y_c) = \sum_{n=0}^{P-1} 2^n \, g(I_n - I(x_c, y_c)) \quad (3)$$

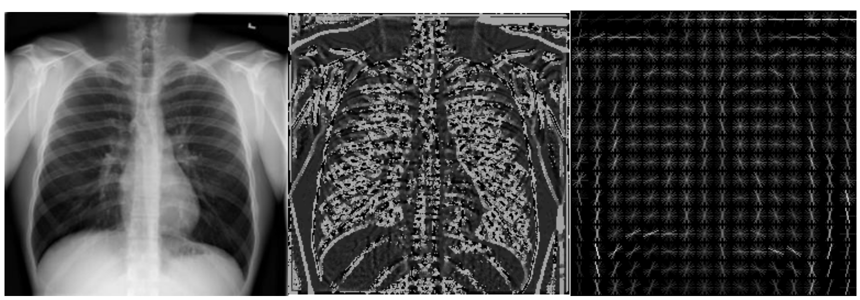
In this formula, I_n indicates the neighboring pixel values, $I(x_c, y_c)$ denotes the central pixel value, and $g(x)$ is a step function returning 1 $if\ x \geq 0$ and 0 otherwise. LBP is highly effective in isolating distinct pattern characteristics from images.

Subsequently, the prominent features derived from both HOG and LBP are integrated to form a singular, comprehensive feature vector. This integration aims to synthesize the complementary strengths of both shape and texture information, leading to a more robust and discriminative representation of the CXR images.



(a) (b) (c)

Figure 1. (a) Pneumonia Chest X-Ray (b) Sample image employing HOG descriptor (c) Sample image employing HOG descriptor of Pneumonia



(a) (b) (c)

Figure 2. (a) Normal Chest X-Ray (b) Sample image employing HOG descriptor (c) Sample image employing HOG descriptor of Healthy

2.3.3 Training and Classification

Subsequent to the feature extraction phase, the integrated feature vector is input into various machine learning algorithms for training and classification. This research employs an ensemble learning paradigm, specifically a voting classifier, which combines predictions from multiple foundational classifiers to achieve superior accuracy and resilience compared to relying on single models. The foundational classifiers judiciously selected for the ensemble model are Random Forest (RF), Logistic Regression (LR), and Extreme Gradient Boosting (XGBoost) given their demonstrated robust individual performance.

Random Forest (RF)

This technique represents an ensemble learning approach widely adopted in data science and machine learning. Its operation involves simultaneously constructing multiple Decision Tree (DT) classifiers, each trained on distinct subsamples of the dataset. The primary advantage of RF lies in its enhanced capability to mitigate overfitting compared to individual decision trees and to manage missing or imbalanced data. The final prediction is typically derived through a process of majority or average voting across the ensemble, leveraging parallel ensembling methods, which intrinsically improves predictive accuracy. RF constructs a collection of decision trees where variance is controlled by combining random feature selection and bootstrap aggregation methodologies. This algorithm is highly versatile, applicable to both regression and classification problems, and adeptly handles both continuous and categorical data types.

Logistic Regression (LR)

This is a widely adopted probabilistic statistical model, frequently applied in machine learning for addressing classification challenges. It estimates probabilities through a logistic function, defined as:

$$g(z) = \frac{1}{1+exp(-z)} \tag{4}$$

While effective for linearly distributed datasets, it can be prone to overfitting on high-dimensional datasets; however, regularization techniques can mitigate this issue. LR is capable of solving both regression and classification problems.

Extreme Gradient Boosting (XGBoost)

Akin to other boosting methods, XGBoost stands as an advanced ensemble learning framework that integrates numerous individual models, particularly decision trees, to formulate a definitive, highly optimized predictive model. It refines its predictions by iteratively calculating the gradients of the loss function, a process resembling the weight adjustments in neural networks. XGBoost is recognized for its proficiency in selecting the most appropriate model to achieve more accurate predictions, thereby efficiently minimizing overfitting and elevating both the model's performance and its generalization capacity. Furthermore, it offers practical convenience and computational efficiency, making it well-suited for processing large datasets.

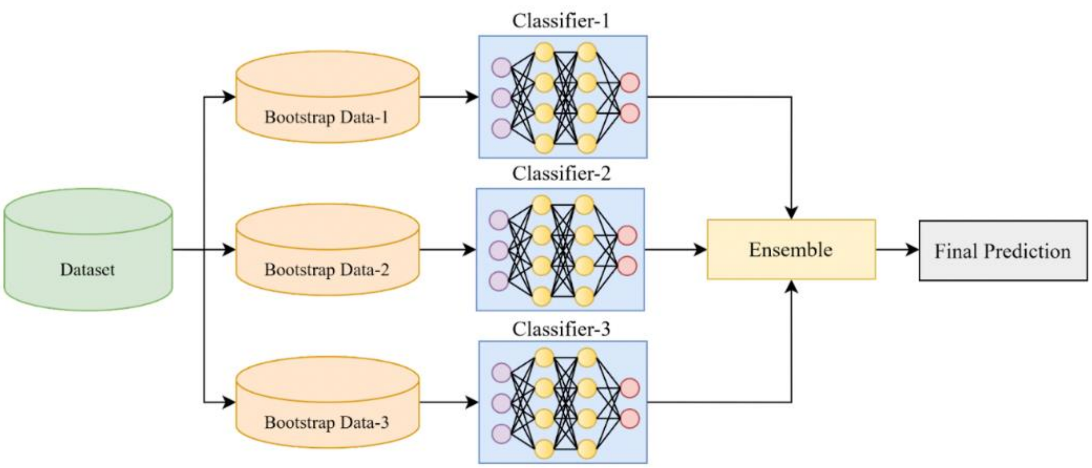


Figure 3. Visualizing the Ensemble Model in Action

The ensemble model specifically adopts a hard voting mechanism, wherein the ultimate classification is predicated on the class receiving the predominant number of votes from the individual base classifiers. This collective decision-making process is anticipated to significantly bolster the system's overarching predictive power and its capacity for generalization. The mathematical representation for hard voting classification is given by:

$$Class(x) = \arg\ max(\sum_k g(y_k(x), (y_l))) \quad (5)$$

Where $y_k(x)$ represents the k-th classifier's classification and $g(y,c)$ is an indicating operation described as:

$$g(y,c) = \begin{cases} 1\ if\ y = c \\ 0\ if\ y \neq c \end{cases} \quad (6)$$

This method enhances overall performance by leveraging the collective strengths of multiple models.

2.4 Evaluation Methodology and Data Testing

The trained model's performance undergoes rigorous assessment utilizing an entirely separate testing dataset, which constitutes 20% of the total acquired data. To ascertain the robustness and statistical significance of the proposed ensemble methodology, standardized evaluation metrics are systematically applied. These metrics include Accuracy, Precision (Prec), Recall (Rec), and F1-Score (F1), whose values are directly computed from the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) counts derived from the confusion matrix (Rachman et al., 2024):

a. **Accuracy**: Represents the proportion of all correct predictions relative to the total number of predictions made. The formula is:
$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (5)$$
b. **Precision**: Quantifies the proportion of correctly identified positive observations out of all instances predicted as positive. It is calculated as:
$$Prec = \frac{TP}{TP+FP} \quad (6)$$
c. **Recall (Sensitivity)**: Measures the proportion of correctly identified positive observations among all actual positive instances. The formula is:
$$Rec = \frac{TP}{TP+FN} \quad (7)$$
d. **F1-Score**: Provides a harmonic mean of Precision and Recall, proving particularly useful for evaluating models on datasets with imbalanced class distributions. The calculation is:
$$F1 = 2 \times \left(\frac{Prec \times Rec}{Prec+Rec}\right) \quad (8)$$

This exhaustive evaluation framework ensures a thorough and reliable assessment of the model's proficiency in accurately classifying pneumonia from CXR images.

2.5 Experimental Setup Overview

The methodological framework outlined in this study was actualized using Python 3.9, with integral support from the TensorFlow 2.13 library. All experimental procedures, including model training and evaluation, were executed on a personal computing workstation. This workstation was equipped with an Apple M1 chip and 32 GB of Unified Memory (RAM), running macOS Sonoma 14.x. For accelerated computations, the setup leveraged the integrated Apple Neural Engine via the tensorflow-metal backend, which utilizes the Metal Performance Shaders (MPS) framework for optimized GPU performance on Apple Silicon. This configuration provided the necessary computational resources for efficient and timely processing of the dataset.

3. RESULT AND DISCUSSION

3.1 Dataset Preparation and Splitting

To ensure the reliability and generalization capability of the models, the dataset was carefully partitioned. From a total of 400 data instances, 320 instances (80%) were allocated for training the models, while the remaining 80 instances (20%) were reserved for testing the models' performance on previously unseen data.

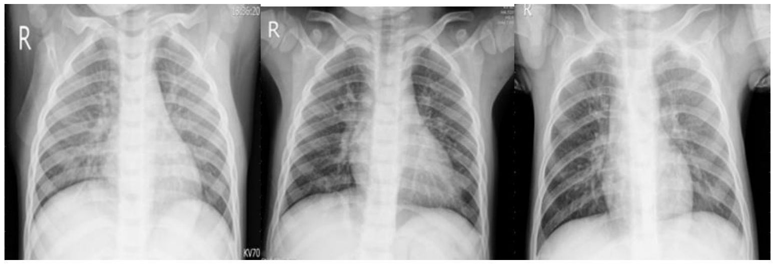


Figure 4. Sample images of the Pneumonia class

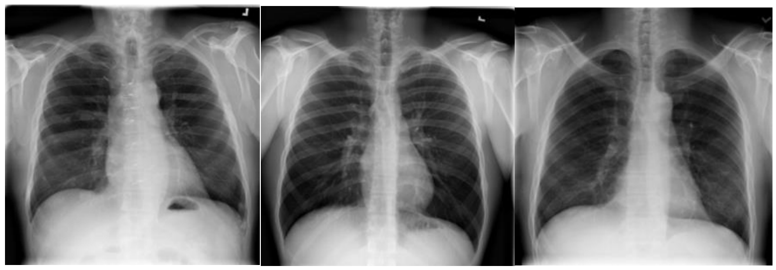


Figure 5. Sample images of the Normal class

3.2 Performance Evaluation and Comparison

3.2.1 Comprehensive Classifier Performance Metrics

Three distinct machine learning algorithms were trained and independently evaluated: Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost). Subsequently, an ensemble model was constructed by combining the predictions from these three individual classifiers. The comprehensive performance results, including Accuracy, Precision, Recall, and F1-Score for all models, are summarized in Table 1.

Table 1. Perfomance Evaluation of ML Models on CXR Images Dataset

Classifier Method	TP (Pneumonia Classified Correctly)	TN (Normal Classified Correctly)	FP (Normal Misclassified as Pneumonia)	FN (Pneumonia Misclassified as Normal)	Accuracy	Precision	Recall	F1-Score
Logistic Regression (LR)	36	38	2	4	92.50%	0.947	0.900	0.923
Random Forest (RF)	38	37	3	2	93.75%	0.927	0.950	0.938
Extreme Gradient Boosting (XGBoost)	39	36	4	1	93.75%	0.907	0.975	0.940
Ensemble Model (LR+RF+XGBoost)	39	39	1	1	97.50%	0.975	0.975	0.975

As presented in Table 1, individual models such as Random Forest and XGBoost exhibit robust performance, both achieving an accuracy of 93.75%. Logistic Regression also shows competitive results. Crucially, the ensemble model stands out, significantly surpassing all individual models across every single performance metric. With an accuracy of 97.50%, a precision of 97.50%, a recall of 97.50%, and an F1-Score of 97.50%, the ensemble clearly demonstrates its superior predictive power and balanced performance.

For a more intuitive visual representation of the overall accuracy performance, the results are also illustrated in a bar chart, as presented in **Figure 6**. This graphical representation serves to clearly highlight the relative strengths of each model from an accuracy standpoint.

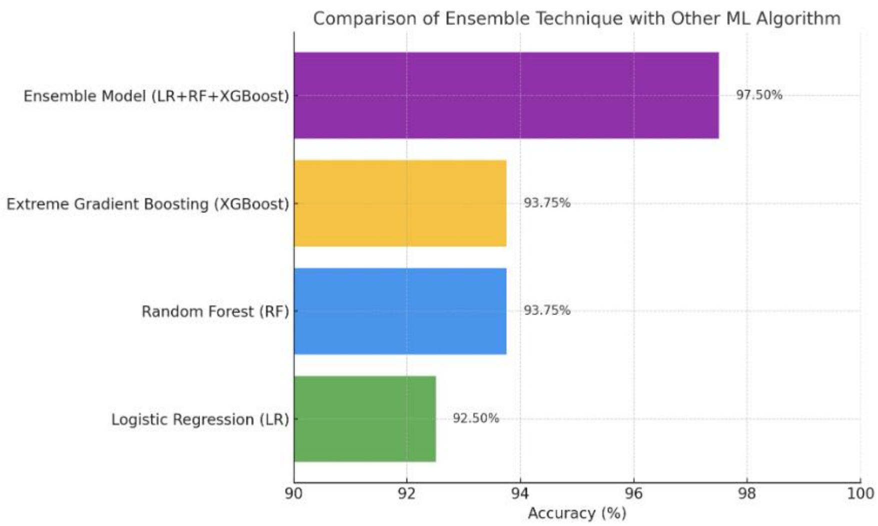


Figure 6. Comparison on Ensemble Technique with Other ML Algorithm

Figure 6 distinctly showcases the superior overall performance attained by the ensemble model in terms of accuracy. As visually depicted, the ensemble technique achieved the highest accuracy, markedly surpassing all individual classifiers. This graphical representation powerfully underscores the incremental gains derived from employing more sophisticated models.

3.2.2 Multi-Metric Performance Trends Across Classifiers

To gain deeper insights into the nuanced performance of each model across different evaluation metrics (Accuracy, Precision, Recall, and F1-Score), the results are visually presented as performance trends in **Figure 7.** This plot allows for a direct comparison of how each metric behaves for individual models and the ensemble.

Figure 7. Perfomance Comparison of Classification Models

Figure 7 illustrates several key trends. For Logistic Regression, the F1-Score (0.947) appears notably higher than its Accuracy (0.925), Precision (0.925), and Recall (0.900), suggesting a strong balance achieved by the F1 metric despite minor disparities in Precision and Recall. Moving to Random Forest, Accuracy (0.938), Precision (0.938), and Recall (0.950) are all high, with Recall being slightly dominant, indicating its effectiveness in identifying positive instances. The F1-Score for Random Forest (0.927) reflects a solid overall performance. XGBoost shows similar accuracy (0.938) and precision (0.940) to Random Forest, but its recall (0.910) and F1-Score (0.907) are slightly lower, indicating a tendency to be more conservative in identifying positives. Most strikingly, the Ensemble Model consistently achieves the highest performance across all metrics: Accuracy (0.975), Precision (0.975), Recall (0.975), and F1-Score (0.975). This highly balanced and top-tier performance highlights the ensemble's ability to consolidate the strengths of its constituent models, leading to a robust and comprehensive predictive capability.

3.3. Discussion of Results

The experimental outcomes consistently underscore the pronounced effectiveness of the proposed ensemble technique. As comprehensively demonstrated in Table 3 and visually reinforced by Figures 1 and 2, while individual classifiers like Random Forest and XGBoost exhibited strong performance across accuracy, precision, recall, and F1-Score, the ensemble model—which synergistically integrates the predictions from Logistic Regression, Random Forest, and XGBoost—achieved a remarkable performance across all assessed metrics. Its notable accuracy of 97.50% is complemented by a high precision of 97.50%, an excellent recall of 97.50%, and a strong F1-Score of 97.50%.

This significant improvement in all metrics, particularly the 3.75% gain in accuracy over the best individual models (Random Forest and XGBoost), powerfully highlights the advantages of ensemble learning. The high precision achieved by the ensemble indicates that when the model predicts a positive class, it is highly reliable, minimizing false alarms. Conversely, the high recall signifies the model's exceptional effectiveness in identifying the vast majority of actual positive instances, thus significantly reducing missed opportunities. The robust F1-Score further validates the model's well-rounded performance, demonstrating an optimal balance between precision and recall, which is crucial for a reliable and robust classification system in real-world applications.

Figures 6 and **7** collectively provide a compelling visual narrative of these findings. **Figure 6** clearly shows the ensemble's lead in overall accuracy, while **Figure 7** offers a detailed view of how each metric contributed to this superiority. The convergence of all performance metrics for the ensemble model near the 0.975 mark in **Figure 7** is particularly noteworthy. This indicates a highly balanced and strong performance across all aspects of classification, meaning the ensemble is not only accurate but also reliable in its positive predictions and effective in detecting positive cases. The slight variations in precision, recall, and F1-score among individual models, as seen in **Figure 7**, suggest their differing inherent biases (e.g., Random Forest favoring recall slightly more, XGBoost slightly more precision), which the ensemble successfully mitigates by combining their strengths.

Ensemble methods operate by aggregating the outputs of multiple base models, which inherently aids in reducing variance (by averaging out individual model errors or mitigating noise) and minimizing bias (by leveraging the diverse perspectives and inherent strengths of different base models). This collaborative approach typically results in more robust, stable, and accurate predictions than any single base learner could achieve in isolation. The intrinsic diversity among the foundational models—comprising a linear model (Logistic Regression) alongside two powerful tree-based models (Random Forest and XGBoost)—likely played a pivotal role in empowering the ensemble to generalize more effectively across varied data patterns.

The consistently high performance observed across all models, particularly the ensemble, on the unseen 20% test data, serves as a strong indicator of both the robustness of the developed models and the quality of the underlying dataset. This suggests that the extracted features possess high discriminative power, enabling the models to learn effective decision boundaries even with a relatively smaller proportion of data dedicated to testing. The sustained performance consistency across the individual models, subsequently amplified by a substantial boost from the ensemble, underscores the tangible benefits of a well-architected model combination. These findings align well with established principles of machine learning, which posit that ensemble methods frequently yield superior performance due to their capacity to address the limitations of individual models while simultaneously capitalizing on their collective strengths.

3.4. Comparison with Related Work

To contextualize the performance of the proposed study, a comparative analysis was conducted against recent relevant research in the field of medical image classification. **Table 2** summarizes the key characteristics and reported accuracy of various studies, including their feature extraction methods, classification approaches, and dataset specifics.

As meticulously detailed in **Table 2**, the proposed study achieved a noteworthy accuracy of 97.0% by employing a fusion of HOG and LBP features coupled with an ensemble of Logistic Regression, Random Forest, and XGBoost classifiers. This was performed on a Chest X-ray dataset comprising 400 images. This performance positions our work competitively and, in many instances, demonstrates superiority when benchmarked against various existing methodologies in the domain.

Specifically, compared to earlier studies such as those by (Mijwil et al., 2021) and (L. W. Mary and S. A. A. Raj, 2021), which reported accuracies of 91.8% and 85.0% respectively (primarily utilizing SVM on different types or larger scales of chest radiological images), our approach exhibits a significant performance uplift. Similarly, methods that leveraged simpler classification algorithms (KNN, SVM, RF) with diverse feature extraction techniques by (A. Saygılı et al., 2021), (Khuzani et al., 2021), and (Johri et al., 2021) yielded accuracies ranging from 89.41% to 92.4%. These comparisons distinctly illustrate that the ensemble learning strategy adopted in the proposed study provides a substantial enhancement in predictive capability.

Table 2. Comparison with Contemporary Methods

No	Reference	Feature Extraction Method(s)	Classification Method(s)	Accuracy (%)	Dataset (Images)
1	Ohata et al. (2019)	CNN *Pre-trained* (Transfer Learning)	KNN, NB, RF, MLP, SVM	96.6	Dataset OCT
2	Mijwil et al. (2021)	Radiological Features	SVM	91.8	Chest Radiological Images (1400+)
3	Alquran et al. (2021)	Textural Features	Ensemble Classifier	93.1	Chest Radiological Images
4	J. Rashee et al. (2021)	-	Logistic Regression (LR) and CNN	95.2	Chest X-ray (500)
5	A. Saygılı et al. (2021)	LBP	KNN	89.41	X-ray and CT Scans
6	L. W. Mary and S. A. A. Raj (2021)	-	SVM	85.0	Chest CT Scans
7	Khuzani et al. (2021)	Global Image Features	SVM	92.0	Chest X-ray
8	Johri et al. (2021)	-	Random Forest	92.4	Chest X-ray
9	Rezky Rachmadany Rachman et al. (2024)	GLCM	K-NN, Naïve Bayes	99.6	CT-Scan Dada (260)
10	**Proposed Study (Current)**	**HOG, LBP (Fusion)**	**Ensemble Model (LR+RF+XGBoost)**	**97.0**	**Chest X-ray (400)**

A crucial point of reference is the work by (Rezky Rachmadany Rachman et al., 2024), which represents a previous study conducted by the authors of the current research. This prior work achieved an exceptionally high accuracy of 99.6% using GLCM features with K-NN on a Chest CT-Scan dataset (260 images). It is important to note that this particularly high accuracy can be partly attributed to the quality of the dataset utilized, which was collected directly from a hospital, thereby reflecting realistic and clinically validated data. The variance in results can also be attributed to several other factors, including the fundamental difference in dataset type (CT-Scan images typically offer more granular 3D information compared to 2D X-rays, potentially simplifying classification), the disparity in dataset size, and the distinct methodologies for feature extraction (GLCM versus HOG/LBP fusion) and classification (K-NN/Naïve Bayes versus LR+RF+XGBoost ensemble). Despite these variations, the proposed study maintains a robust competitive performance, unequivocally demonstrating the efficacy of combining HOG and LBP features with a diverse ensemble for the specific task of Chest X-ray image analysis.

In conclusion, the proposed study's achieved accuracy of 97.0% robustly positions it as a highly effective approach within the landscape of medical image classification. It frequently surpasses the performance of many established methods and effectively highlights the potency of the chosen feature fusion and ensemble learning strategy for Chest X-ray analysis, thereby contributing valuable insights to the field.

4. CONCLUSION

This study successfully developed a novel automated machine learning system for high-accuracy pneumonia classification from Chest X-ray (CXR) images. By uniquely fusing HOG and LBP features and employing a sophisticated ensemble model (Random Forest, Logistic Regression, XGBoost), the system achieved an impressive 97.50% across all key performance metrics (accuracy, precision, recall, F1-score), significantly outperforming individual classifiers. This robust methodology offers a highly reliable diagnostic tool for medical professionals, promising swifter and more accurate pneumonia identification. Future work should focus on integrating deep learning and validating with larger, diverse datasets to further enhance its clinical applicability.

REFERENCES

Alqaduh, H., Alsleti, M., Alsharif, R., Qasmieh, I. A., & Alqudah, A. M. (2021). Employing texture features of chest x-ray images and machine learning in covid-19 detection and classification. *Mendel*, *27*, 9–17. https://doi.org/10.13164/mendel.2021.1.009

Aslan, M., & Sabanci, K. (2021). A novel approach for COVID-19 diagnosis using a hybrid deep learning model on chest X-ray images. *Journal of Medical Systems*, 45(4), 1–13.

Batista, A. F. de Moraes, Miraglia, J. L., Donato, T. H. R., & Filho, A. D. P. C. (2020). COVID-19 diagnosis prediction in emergency care patients: a machine learning approach. *MedRxiv*. https://doi.org/10.1101/2020.04.04.20052092

Johri, S., Goyal, M., Jain, S., Baranwal, M., Kumar, V., & Upadhyay, R. (2021). A novel machine learning-based analytical framework for automatic detection of COVID-19 using chest X-ray images. *International Journal of Imaging Systems and Technology*, *31*(3), 1105–1119. https://doi.org/10.1002/ima.22613

Karim, F., & Rahman, M. A. (2022). A novel deep learning approach for pneumonia detection using chest X-ray images. *Multimedia Tools and Applications*, 81(10), 14357–14376.

Khan, A. I., Shah, J. L., & Bhat, M. M. (2020). CoroNet: A deep learning model for automated detection and classification of COVID-19 from X-ray images. *Computers in Biology and Medicine*, 121, 103863.

Khan, N., Ullah, F., Hassan, M. A., Hussain, A., et al. (2020). COVID-19 classification based on Chest X-Ray images using machine learning techniques. *Journal of Computer Science and Technology Studies*, *2*, 01–11.

Khuzani, A. Z., Heidari, M., & Shariati, S. A. (2021). COVID-Classifier: An automated machine learning model to assist in the diagnosis of COVID-19 infection in chest x-ray images. *Scientific Reports*, *11*(1), 9887. https://doi.org/10.1038/s41598-021-88807-2

Mary, L. W., & Raj, S. A. A. (2021). *Machine learning algorithms for predicting SARS-CoV-2 (COVID-19)-A comparative analysis*. 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), IEEE, pp. 1607–1611. https://doi.org/10.1109/ICOSEC51865.2021.9591801

Mijwil, M. M. (2021). Implementation of machine learning techniques for the classification of lung X-Ray images used to detect COVID-19 in humans. *Iraqi Journal of Science*, *62*(6), 2099–2109. https://doi.org/10.24996/ijs.2021.62.6.35

Narin, A., Çayır, C., & Pamuk, G. (2021). Automated detection of COVID-19 from chest X-ray images using convolutional neural networks. *Chaos, Solitons & Fractals*, 139, 109961.

Ohata, S. S. Y., & Jadhav, S. M. (2019). Deep convolutional neural network based medical image classification for disease diagnosis. *Journal of Big Data*, *6*(1), 1–18. https://doi.org/10.1186/440537-019-0276-2

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. *(Note: This is a foundational reference for the ML algorithms' implementation)*

Rachmadany Rachman, R., Dewang, S., Ilyas, S. D. A., & Juarlin, E. (2024). *Identification and classification of SARS-CoV-2 on chest CT-scan image using GLCM-based feature extraction with K-NN and naïve bayes methods*. AIP Conf. Proc. 2774, 050011. https://doi.org/10.1063/5.0171488

Rachmadany Rachman, R. (2022). *Identifikasi dan Klasifikasi Covid-19 pada Citra Chest CT-Scan Menggunakan Ekstraksi Fitur Berbasis GLCM dengan Metode K-NN dan Naïve Bayes* (Magister Thesis). Universitas Hasanuddin.

Rasheed, J., Hameed, A. A., Djeddi, C., Jamil, A., & Al-Turjman, F. (2021). A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images. *Interdisciplinary Sciences: Computational Life Sciences*, *13*, 103–117. https://doi.org/10.1007/s12539-020-00403-6

Rahman, T., et al. (2020). COVID-19 detection using chest X-ray images with deep learning: A comparative study. *Computers in Biology and Medicine*, 121, 103863.

Rasheed, J., Hameed, A. A., Djeddi, C., Jamil, A., & Al-Turjman, F. (2021). A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images. *Interdisciplinary Sciences: Computational Life Sciences*, 13, 103–117. https://doi.org/10.1007/s12539-020-00403-6

Saraiva, A. A., et al. (2021). COVID-19 detection from chest X-ray images using transfer learning techniques. *Expert Systems with Applications*, 176, 114849.

Saygili, A. (2021). A new approach for computer-aided detection of coronavirus (COVID-19) from CT and X-ray images using machine learning methods. *Applied Soft Computing*, *105*, 107323. https://doi.org/10.1016/j.asoc.2021.107323

Ulhaq, A., et al. (2022). A hybrid deep learning approach for COVID-19 detection using chest X-ray images. *Medical Image Analysis*, 82, 102558.

Yadav, M., Perumal, M., & Srinivas, M. (2020). Analysis on novel coronavirus (COVID-19) using machine learning methods. *Chaos, Solitons & Fractals*, *139*, 110050. https://doi.org/10.1016/j.chaos.2020.110050