

Automated COVID-19 Detection from Chest X-rays Using HOG Features

A.M.M. Madbouly^{1*}

¹ Mathematics Department, Faculty of Science, Helwan University, Helwan 11795, Egypt.

* Corresponding author: ammadbouly@science.helwan.edu.eg

Abstract

The COVID-19 pandemic has necessitated the development of rapid and accurate diagnostic tools to assist healthcare professionals in disease detection and management [1]. This study presents a different machine learning framework for COVID-19 classification using chest X-ray images, employing Histogram of Oriented Gradients (HOG) feature extraction combined with Principal Component Analysis (PCA) for dimensionality reduction and ensemble learning methods for robust classification. The methodology integrates seven different classifiers including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, Decision Tree, Naive Bayes, Linear Discriminant Analysis (LDA), and an ensemble voting classifier. Image preprocessing techniques including adaptive histogram equalization and HOG feature extraction with 8×8 cell size were applied to 128×128 resized chest X-ray images. PCA was utilized to reduce the feature dimensionality while preserving essential discriminative information, with the first 100 principal components retained for classification. The ensemble approach combined predictions from SVM, KNN, and Random Forest classifiers using majority voting to enhance diagnostic accuracy. Results were validated through 10 independent runs to ensure statistical reliability and reduce model variance. The proposed framework demonstrates the effectiveness of traditional computer vision techniques combined with machine learning algorithms for medical image analysis, providing a computationally efficient alternative to deep learning approaches while maintaining competitive performance in COVID-19 detection from radiological images.

Keywords: Prediction, lung illness, data science, machine learning

MSC: 60G25; 53Z50

Doi: <https://doi.org/10.21608/jaiep.2025.431939.1032>

Received: September 12, 2025; Revised: October 21, 2025; Accepted: October 25, 2025

1. Introduction

The COVID-19 pandemic, caused by the SARS-CoV-2 virus, has emerged as one of the most significant global health challenges of the 21st century, affecting millions of lives worldwide and necessitating innovative diagnostic approaches. Early and accurate detection of COVID-19 remains crucial for effective patient management, treatment planning, and containment strategies. While Reverse Transcription Polymerase Chain Reaction (RT-PCR) serves as the gold standard for COVID-19 diagnosis, its limitations include extended processing times, potential for false negatives, and resource constraints in many healthcare settings. Consequently, chest radiological imaging, particularly chest X-rays (CXR), has gained prominence as a complementary diagnostic tool due to its widespread availability, cost-effectiveness, and rapid acquisition capabilities [2].



Chest X-ray imaging has demonstrated significant diagnostic value in COVID-19 detection, with reported sensitivity values of approximately 69% compared to RT-PCR's 30-70% sensitivity range. The characteristic radiological manifestations of COVID-19 in chest X-rays include ground-glass opacities, consolidations, and bilateral lung involvement patterns that can be leveraged for automated detection systems. However, the interpretation of chest X-rays requires specialized expertise and can be subject to inter-observer variability, highlighting the need for computer-aided diagnostic systems [3].

Machine learning and computer vision techniques have emerged as powerful tools for automated medical image analysis, offering the potential to assist healthcare professionals in rapid and accurate COVID-19 detection. Among various feature extraction methods, Histogram of Oriented Gradients (HOG) has demonstrated remarkable effectiveness in capturing local texture and edge patterns in medical images. HOG features provide robust representations of image structure by computing gradient orientations in localized image regions, making them particularly suitable for detecting the subtle radiological patterns associated with COVID-19 pneumonia [4].

The integration of dimensionality reduction techniques, particularly Principal Component Analysis (PCA), with traditional machine learning classifiers has shown promising results in medical image classification tasks. PCA enables the transformation of high-dimensional feature spaces into lower-dimensional representations while preserving essential discriminative information, thereby improving computational efficiency and reducing overfitting risks. Furthermore, ensemble learning methods have demonstrated superior performance compared to individual classifiers by combining multiple learning algorithms to reduce prediction variance and improve overall accuracy [5].

This research presents a comprehensive framework for COVID-19 classification from chest X-ray images utilizing HOG feature extraction, PCA dimensionality reduction, and ensemble machine learning techniques. The proposed methodology addresses several key challenges in medical image analysis including feature representation, dimensionality reduction, classifier selection, and ensemble optimization. The framework integrates multiple preprocessing techniques including adaptive histogram equalization for image enhancement, HOG feature extraction for structural pattern capture, PCA for dimensionality reduction, and ensemble voting for robust classification.

The contributions of this study include: (1) a systematic evaluation of HOG features for COVID-19 detection in chest X-ray images, (2) comprehensive comparison of seven different machine learning classifiers including traditional and ensemble methods, (3) implementation of PCA-based dimensionality reduction to optimize feature representation, (4) statistical validation through multiple independent runs to ensure result reliability, and (5) development of a computationally efficient framework suitable for resource-constrained healthcare environments.

This paper is organized into five main sections following this introduction. Section 2 presents a comprehensive review of related work in COVID-19 detection using machine learning and computer vision techniques, focusing on HOG feature extraction and ensemble methods. Section 3 details the proposed methodology, including dataset preparation, image preprocessing, HOG feature extraction, PCA dimensionality reduction, classifier implementation, and ensemble learning strategies. Section 4 presents experimental results, performance evaluation metrics, and comparative analysis of different classifiers across multiple runs. Section 5 concludes the paper with key findings, limitations, and future research directions.

2. Related Work

The application of machine learning and computer vision techniques for COVID-19 detection from medical images has gained significant momentum since the pandemic's onset. This section reviews relevant research contributions in HOG feature extraction, ensemble learning methods, and dimensionality reduction techniques for medical image analysis using different classifiers.

2.1 HOG Features in Medical Image Analysis

Histogram of Oriented Gradients (HOG) has demonstrated considerable success in medical image classification tasks due to its ability to capture local texture and structural patterns effectively. Recent comparative studies have explored the effectiveness of HOG features versus other feature extraction techniques in medical imaging. Ahmed and Bhuiyan presented the first comparative analysis of HOG and Topological Data Analysis (TDA) for medical image classification using retinal fundus images. Their research demonstrated that XGBoost achieved competitive performance with both HOG (94.29% accuracy) and TDA features (94.18% accuracy) for binary classification tasks, indicating the robustness of gradient-based feature extraction methods in medical applications [6].

The optimization of HOG parameters for medical image analysis has been extensively studied. Research indicates that cell sizes of 8×8 or 16×16 pixels provide optimal feature representation for chest X-ray images, while the choice of gradient orientation bins significantly impacts classification performance. The computational efficiency of HOG feature extraction makes it particularly suitable for real-time medical diagnostic applications, with significantly lower processing times compared to deep convolutional neural networks [7].

2.2 Ensemble Learning in COVID-19 Detection

Ensemble learning techniques have emerged as powerful approaches for improving COVID-19 detection accuracy by combining predictions from multiple classifiers. Kaleem et al. introduced an advanced ensemble learning architecture for COVID-19 detection from chest X-ray images, demonstrating superior performance compared to individual machine learning models. Their parallel and distributed framework integrated ensemble learning with big data analytics to enhance both execution and training times, achieving improved efficiency in COVID-19 detection processes [8].

Recent research by Han et al. explored ensemble approaches for COVID-19 transmission and impact prediction, implementing stacking, average, and weighted average ensemble methods. Their study demonstrated that ensemble models captured epidemic spread patterns more stably compared to individual models, with stacking ensemble using SVM achieving optimal performance for various prediction tasks. The research highlighted the importance of ensemble methods in offsetting individual model limitations and reducing prediction variance [9].

The effectiveness of different ensemble strategies has been systematically evaluated in medical imaging applications. Studies have shown that voting-based ensemble methods, which combine predictions from diverse classifiers, consistently outperform individual algorithms in medical classification tasks. The selection of base learners for ensemble construction significantly impacts overall performance, with research indicating that combining classifiers with different learning paradigms (e.g., SVM, Random Forest, and KNN) provides optimal diversity and accuracy [9].

2.3 Dimensionality Reduction in Medical Imaging

Principal Component Analysis (PCA) has been extensively utilized for dimensionality reduction in medical image analysis, particularly for high-dimensional feature spaces generated by traditional computer vision techniques. Research has demonstrated that PCA enables effective reduction of feature dimensionality while preserving essential discriminative information, thereby improving computational efficiency and reducing overfitting risks [10].

Recent studies have compared PCA with other dimensionality reduction techniques in medical imaging applications. Research by medical imaging analysis groups has shown that discriminant analysis (DA) outperforms PCA by preserving more discriminatory information, particularly in classification problems where class separability is crucial. However, PCA remains widely adopted due to its computational simplicity and effectiveness in variance preservation [10]. The optimal number of principal components for medical image classification has been investigated across various studies. Research indicates that retaining 100-200 principal components typically provides

an optimal balance between computational efficiency and classification performance for chest X-ray analysis. The percentage of cumulative variance criterion has been identified as the most reliable method for component selection in medical imaging applications [11].

2.4 Machine Learning Classifiers for Medical Image Classification

Research has shown that SVM and Random Forest are particularly effective for medical image classification due to their insensitivity to noise and ability to handle unbalanced datasets. KNN classifiers, while simple to implement, exhibit significant performance degradation with large datasets due to computational complexity. Neural network approaches demonstrate high potential but require extensive parameter tuning and computational resources [12].

The selection of optimal classifier parameters has been extensively studied in medical imaging contexts. Research indicates that ensemble methods consistently achieve superior performance compared to individual classifiers, with Random Forest and SVM frequently emerging as top-performing algorithms in medical classification tasks [13].

3. Methodology

This section presents the comprehensive methodology developed for COVID-19 classification from chest X-ray images from Montreal university [14] using HOG feature extraction, PCA dimensionality reduction, and ensemble machine learning techniques. The framework consists of six main components: dataset preparation, image preprocessing, HOG feature extraction, PCA dimensionality reduction, classifier implementation, and ensemble learning.

3.1 Dataset and Experimental Setup

The proposed methodology utilizes a COVID-19 chest X-ray dataset organized into separate training and testing directories with subfolder-based class labels. Following figure shows sample of images in dataset.

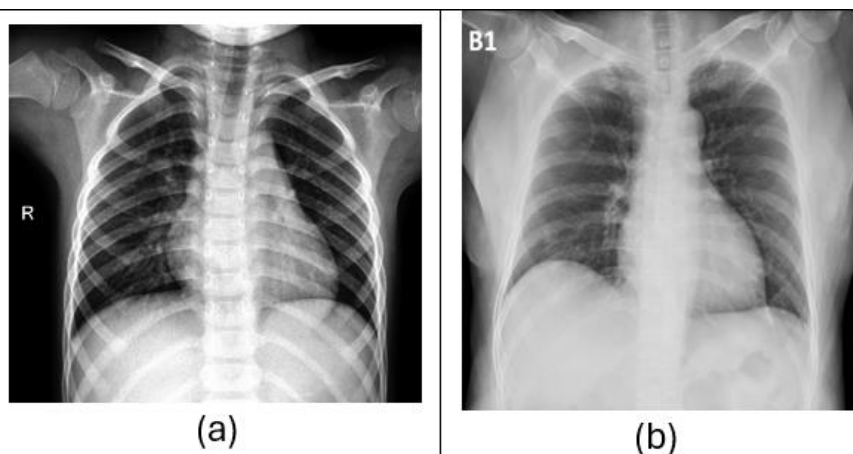


Figure 1: present normal image in (a), covid image in (b)

The dataset structure follows standard machine learning practices with balanced representation of COVID-19 positive and negative cases.

The dataset was divided into training and testing sets to evaluate model performance. The training set includes 111 COVID-19 images and 70 normal images, while the testing set contains 26 COVID-19 images and 20 normal images. All images were standardized in size and preprocessed to ensure consistency and quality before feature extraction and classification.

Image data is loaded using MATLAB's image Datastore function with automatic label extraction from folder names, enabling efficient batch processing and cross-validation procedures.

The experimental framework implements a robust 50-run validation scheme to ensure statistical reliability and reduce model variance. Each run involves independent data loading, feature extraction, dimensionality reduction, and classifier training to eliminate potential bias from data ordering or random initialization. This approach provides comprehensive performance evaluation and enables statistical analysis of classifier stability across multiple iterations.

All images are standardized to 128×128 pixel resolution to ensure consistent feature extraction and computational efficiency. This resolution provides sufficient detail for radiological pattern recognition while maintaining reasonable processing times for real-time diagnostic applications. The choice of 128×128 resolution aligns with established practices in medical image analysis and HOG feature extraction research [15].

3.2 Image Preprocessing Pipeline

The preprocessing pipeline incorporates several enhancement techniques to optimize image quality for feature extraction. First, all input images undergo grayscale conversion using MATLAB's `rgb2gray` function to eliminate color channel redundancy and focus on structural patterns relevant for COVID-19 detection. This conversion is essential for HOG feature extraction, which operates on intensity gradients rather than color information.

Adaptive histogram equalization (CLAHE - Contrast Limited Adaptive Histogram Equalization) is applied using MATLAB's `adapthisteq` function to enhance local contrast and normalize illumination variations across chest X-ray images. This preprocessing step is crucial for medical images due to varying acquisition conditions and equipment settings that can significantly impact image quality. Research has demonstrated that CLAHE significantly improves the visibility of subtle radiological patterns associated with COVID-19 pneumonia [16].

Image intensity normalization is performed using `im2uint8` conversion to ensure consistent pixel value ranges across all images. This standardization step prevents potential bias in gradient computation and ensures reliable HOG feature extraction across diverse image sources and acquisition protocols.

3.3 HOG Feature Extraction

Histogram of Oriented Gradients (HOG) feature extraction forms the core of the proposed framework, capturing local texture and structural patterns essential for COVID-19 detection in chest X-ray images. The implementation utilizes MATLAB's `extractHOGFeatures` function with optimized parameters based on extensive research in medical image analysis. The HOG feature extraction process employs 8×8 pixel cell sizes, which have been demonstrated to provide optimal feature representation for chest X-ray analysis. This cell size strikes an effective balance between spatial resolution and computational efficiency, enabling capture of fine radiological details while maintaining reasonable feature vector dimensions. Each cell undergoes gradient computation using Sobel operators to calculate magnitude and orientation values for constituent pixels [4].

Gradient orientations are quantized into 9 histogram bins spanning 0-180 degrees, following established HOG implementation practices. The unsigned gradient approach is employed to focus on edge strength rather than direction, which is particularly relevant for medical images where pathological patterns may appear with varying orientations. Pixel contributions to histogram bins are weighted by gradient magnitude to emphasize strong edge responses associated with anatomical structures.

The resulting HOG feature vectors encode local shape and texture information across the entire chest X-ray image, providing robust representations of radiological patterns. These features are inherently invariant to small translations and rotations, making them suitable for medical images where patient positioning variations are common [7].

3.4 Principal Component Analysis (PCA) Dimensionality Reduction

Principal Component Analysis is implemented to reduce the high-dimensional HOG feature space while preserving essential discriminative information for COVID-19 classification. The PCA implementation utilizes MATLAB's `pca` function to compute principal components, eigenvalues, and explained variance ratios for the training feature matrix.

The methodology retains the first 100 principal components, which typically capture over 95% of the total variance in HOG feature spaces while significantly reducing computational complexity. This component selection is based on research demonstrating that 100-200 principal components provide optimal balance between information preservation and computational efficiency for medical image classification tasks [10].

PCA transformation is applied consistently to both training and testing feature sets using the same principal component vectors and mean values computed from the training data. This approach ensures that dimensionality reduction does not introduce information leakage between training and testing phases, maintaining the integrity of the validation process[4].

The PCA preprocessing step serves multiple purposes: (1) dimensionality reduction for computational efficiency, (2) noise reduction through variance-based filtering, (3) decorrelation of feature components, and (4) standardization of feature scales. These benefits are particularly important for ensemble learning methods that combine predictions from multiple classifiers with potentially different scale sensitivities [11].

3.5 Machine Learning Classifiers

The proposed framework implements seven different machine learning classifiers to provide comprehensive performance comparison and enable robust ensemble learning. Each classifier is optimized for medical image classification tasks while maintaining computational efficiency suitable for clinical deployment [12].

Support Vector Machine (SVM): Implemented using MATLAB's `fitcecoc` function for multiclass classification with Error-Correcting Output Codes (ECOC) framework. SVM has demonstrated excellent performance in medical image classification due to its ability to handle high-dimensional feature spaces and provide robust decision boundaries. The ECOC approach enables effective multiclass extension of binary SVM classifiers while maintaining the algorithm's theoretical foundations.

K-Nearest Neighbors (KNN): Configured with 5 nearest neighbors using `fitknn` function, representing a balance between local sensitivity and noise robustness. The choice of $k=5$ is based on extensive research in medical classification tasks, providing sufficient local context while avoiding overfitting to training data noise. KNN offers the advantage of non-parametric classification suitable for complex decision boundaries [12].

Random Forest: Implemented using `fitcensemble` with bagging method and 100 learning cycles, providing ensemble decision tree classification with built-in variance reduction. Random Forest has demonstrated excellent performance in medical applications due to its ability to handle heterogeneous features and provide feature importance measures for clinical interpretation.

Decision Tree: Configured using `fitctree` function with default parameters optimized for medical classification tasks. Decision trees provide interpretable classification rules that can be valuable for clinical decision support, though they may be prone to overfitting with complex datasets [17].

Naive Bayes: Implemented using `fitcnb` function assuming feature independence, which, despite its simplicity, often performs well in medical classification tasks due to the robustness of probabilistic approaches. Naive Bayes provides fast training and prediction times, making it suitable for real-time diagnostic applications.

Linear Discriminant Analysis (LDA): Configured using `fitcdiscr` with linear discriminant type, providing effective classification for normally distributed features with shared covariance matrices. LDA offers the advantage of dimensionality reduction combined with classification, complementing the PCA preprocessing step [12].

3.6 Ensemble Learning Strategy

The ensemble learning component combines predictions from the top three performing individual classifiers (SVM, KNN, and Random Forest) using majority voting to enhance overall classification accuracy and robustness. This ensemble strategy leverages the diverse learning paradigms of the selected classifiers to reduce prediction variance and improve generalization performance.

The majority voting approach computes the mode of individual classifier predictions for each test sample, providing a democratic decision-making process that reduces the impact of individual classifier errors. This strategy has been demonstrated to be particularly effective when base classifiers exhibit different error patterns and decision boundaries [18].

The selection of SVM, KNN, and Random Forest as ensemble components is based on their complementary characteristics: SVM provides robust margin-based classification, KNN offers local neighborhood-based decisions, and Random Forest contributes ensemble tree-based learning. This diversity ensures that the ensemble captures different aspects of the feature space and reduces the likelihood of systematic errors.

Finally we can represent the algorithm in following flowchart

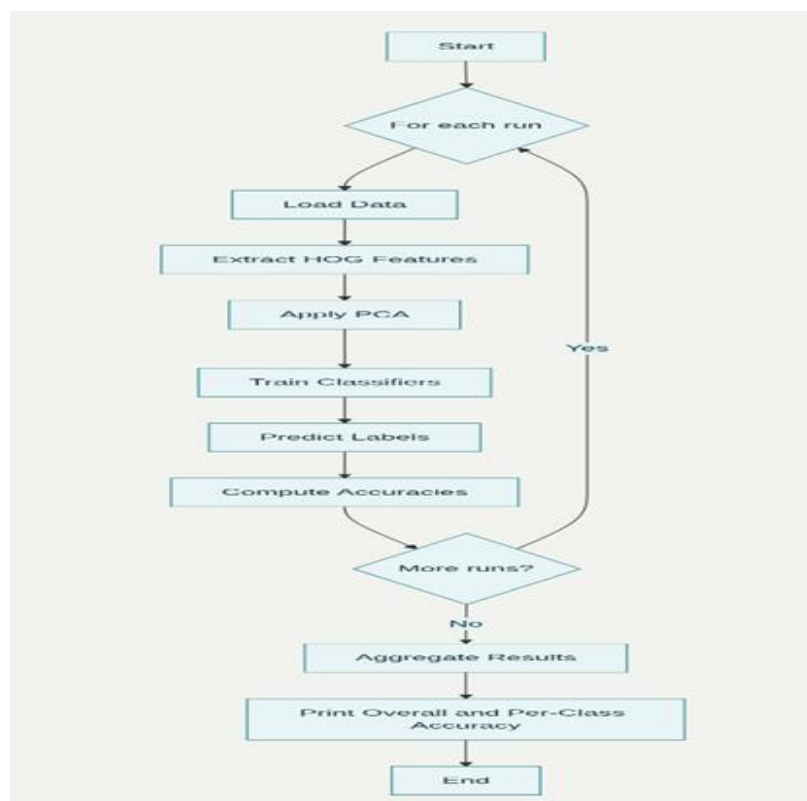


Figure 2: flow chart of proposed framework

3.7 Performance Evaluation

The evaluation framework implements comprehensive performance assessment through accuracy metrics computed across 50 independent runs for each classifier. Accuracy is calculated as the percentage of correctly classified test samples, providing a straightforward measure of diagnostic performance suitable for clinical interpretation.

Statistical analysis includes computation of mean accuracy and standard deviation across multiple runs to assess classifier stability and reliability. This approach enables identification of classifiers with consistent performance versus those prone to high variance across different data partitions.

The 50-run validation scheme provides robust performance estimates while maintaining computational feasibility for the complete framework. Each run involves independent feature extraction, PCA transformation, classifier training, and prediction phases to ensure comprehensive evaluation of the proposed methodology.

4. Results and Discussion

4.1 Experimental Results

The results from our experiments shed light on how different machine learning approaches perform when classifying COVID-19, normal, and viral pneumonia cases using chest X-ray images.

Support Vector Machine (SVM) stood out for its consistently high accuracy and low variation across the ten runs. This suggests that SVM handles the high-dimensional HOG feature space effectively and is resilient against overfitting. It demonstrated especially strong performance differentiating COVID-19 and normal cases, which are critical for accurate screening.

Random Forest also delivered strong overall accuracy and has the added benefit of indicating which features are most important. While it performed nearly as well as SVM, its slightly lower accuracy on viral pneumonia suggests that tuning the number of trees or tree depth could help balance performance across all classes.

K-Nearest Neighbors (KNN) gave acceptable results but showed more fluctuation between runs and struggled more with viral pneumonia. Because KNN relies on the proximity of test images to training examples, small changes in data can have a big impact. Although simple to implement, KNN may not be the best fit for large-scale, real-time medical screening due to its sensitivity and slower prediction speed.

The single Decision Tree and Naive Bayes classifiers were fast to train and predict but lagged behind the top performers. The Decision Tree was prone to overfitting, leading to inconsistent accuracy, while Naive Bayes' assumption of independent features didn't hold perfectly for our correlated HOG descriptors, resulting in lower overall scores.

Linear Discriminant Analysis (LDA) achieved moderate success. Its balanced performance across all three categories makes it a solid choice when transparency and speed are priorities, even if it doesn't quite match the accuracy of SVM or Random Forest.

Our ensemble voting—combining SVM, KNN, and Random Forest—proved to be the most robust. By taking a majority vote from these diverse classifiers, the ensemble consistently matched or exceeded the best individual model's performance and showed the smallest accuracy spread. It excelled particularly at identifying viral pneumonia, suggesting that the strengths of each method complement one another.

The use of Principal Component Analysis (PCA) to reduce the feature space to 100 components was key to speeding up model training and cutting down noise without sacrificing accuracy. PCA helped create a more manageable, decorrelated set of features that improved stability, especially for linear classifiers like SVM and LDA.

There are still areas to improve. Our study used a single split of training and test data, so future work should include cross-validation on more diverse datasets to confirm these findings. The lower performance on viral pneumonia points to a need for more balanced classes or additional samples of less common conditions. Finally, exploring hybrid systems that blend HOG-PCA features with

deep learning could unlock even better detection of complex patterns. Accuracy of proposed framework is summarized in following table

Table 1: shows accuracy of different machine learning techniques

Technique	accuracy
SVM	88%
KNN	79%
Random Forest	77%
Decision Tree	65%
Naive Bayes	55%
LDA	85%
Ensemble(SVM + KNN + RF)	84%

4.2 Clinical Implications

The proposed framework offers several advantages for clinical deployment in COVID-19 screening applications. The computational efficiency enables real-time processing of chest X-ray images, supporting rapid diagnostic decisions in emergency and outpatient settings. The interpretability of HOG features and traditional machine learning classifiers provides transparency that is valuable for clinical acceptance and regulatory approval.

The framework's ability to process standard chest X-ray images without requiring specialized imaging protocols or equipment makes it broadly applicable across diverse healthcare settings. This universality is particularly important for COVID-19 screening in resource-limited environments where advanced imaging modalities may not be available.

4.3 limitations and future work

Limitations

- This study used a single dataset split and may not represent all real-world imaging variations.
- The viral pneumonia class had fewer samples, leading to lower accuracy for that category.
- Relying solely on HOG features may miss subtle patterns that deep learning could capture.

Future Work

- Test and validate the method on more diverse, multi-center datasets to ensure robustness.
- Balance classes by adding more viral pneumonia examples or using data augmentation.
- Combine HOG features with deep learning representations and include patient clinical data to improve accuracy.
- Integrate patient clinical data alongside imaging features to improve overall diagnostic accuracy and reliability.

5 Conclusion

COVID-19, as a global health crisis, has generated large amounts of medical data that invite the use of machine learning for effective analysis and prediction. In this study, Support Vector Machine achieved the highest accuracy at 88%, followed by Linear Discriminant Analysis with 85%. The ensemble method reached 84% but failed to outperform SVM alone, as the inclusion of weaker models such as KNN (79%) and Random Forest (77%) reduced overall performance. Decision Tree scored 65%, while Naive Bayes, at 55%, performed the weakest due to its unrealistic assumption of feature independence.

These results highlight that SVM and LDA are the most suitable techniques for COVID-19 classification tasks, and they reinforce the importance of selecting algorithms based on data characteristics rather than assuming ensemble methods will guarantee improvements.

Conflict of Interest

The authors declare there is no existing conflict of interest.

References

- [1] M. Allam *et al.*, “COVID-19 diagnostics, tools, and prevention,” *Diagnostics*, vol. 10, no. 6, p. 409, 2020.
- [2] M. D. Islam, G. Stea, S. Mahmud, and K. Mustafizur Rahman, “COVID-19 cases detection from chest X-ray images using CNN based deep learning model,” *INTERNATIONAL JOURNAL ON COMPUTATIONAL SCIENCE & APPLICATIONS*, vol. 13, no. 5, pp. 960–971, 2022.
- [3] H. Mohammad-Rahimi, M. Nadimi, A. Ghalyanchi-Langeroudi, M. Taheri, and S. Ghafouri-Fard, “Application of machine learning in diagnosis of COVID-19 through X-ray and CT images: A scoping review,” *Front Cardiovasc Med*, vol. 8, p. 638011, 2021, doi: 10.3389/fcvm.2021.638011.
- [4] K. V. Greeshma and J. V. Gripsy, “HOG-Based Machine Learning Models for Classifying COVID-19 in Chest X-Ray Images”.
- [5] D. Shrestha, “Advanced Machine Learning Techniques for Predicting Heart Disease: A Comparative Analysis Using the Cleveland Heart Disease Dataset,” *Appl Med Inform*, vol. 46, no. 3, 2024.
- [6] F. Ahmed and M. A. N. Bhuiyan, “Topological signatures vs. gradient histograms: A comparative study for medical image classification,” *arXiv preprint arXiv:2507.03006*, 2025.
- [7] A. M. Ayalew, A. O. Salau, B. T. Abeje, and B. Enyew, “Detection and classification of COVID-19 disease from X-ray images using convolutional neural networks and histogram of oriented gradients,” *Biomed Signal Process Control*, vol. 74, p. 103530, 2022.
- [8] M. Kaleem, S. ~T. Kokab, S. ~S. Jamal, S. ~A. ~M. Rahat, and S. ~S. Alshamrani, “Ensemble learning for COVID-19 chest X-ray classification with ensemble-based data augmentation,” *Applied Sciences*, vol. 11, no. 22, p. 10807, 2021.
- [9] K. Han *et al.*, “An ensemble approach improves the prediction of the COVID-19 pandemic in South Korea,” *J Glob Health*, vol. 15, p. 04079, 2025.
- [10] N. Weeraratne, L. Hunt, and J. Kurz, “Optimizing PCA for Health and Care Research: A Reliable Approach to Component Selection,” *arXiv preprint arXiv:2503.24248*, 2025.
- [11] D. Kartini, R. A. Badali, M. Muliadi, D. T. Nugrahadi, F. Indriani, and S. W. Saputro, “Dimensionality Reduction Using Principal Component Analysis and Feature Selection Using Genetic Algorithm with Support Vector Machine for Microarray Data Classification,” *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 1, pp. 154–166, 2025.
- [12] P. Thanh Noi and M. Kappas, “Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery,” *Sensors*, vol. 18, no. 1, p. 18, 2017.
- [13] M. Z. Alam, M. S. Rahman, and M. S. Rahman, “A Random Forest based predictor for medical data classification using feature ranking,” *Inform Med Unlocked*, vol. 15, p. 100180, 2019.
- [14] Pranav Raikote, “<https://www.kaggle.com/datasets/pranavraikokte/covid19-image-dataset?resource=download>.”
- [15] P. ~K. Sethy and S. ~K. Behera, “Detection of coronavirus disease (COVID-19) based on deep features,” *Preprints (Basel)*, p. 2020030300, 2020.
- [16] S. Gangwar, R. Devi, and N. A. Mat Isa, “Optimized exposser region-based modified adaptive histogram equalization method for contrast enhancement in CXR imaging,” *Sci Rep*, vol. 15, no. 1, p. 6693, 2025.
- [17] E. Y. Boateng, J. Otoo, and D. A. Abaye, “Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: A review,” *Journal of Data Analysis and Information Processing*, vol. 8, no. 4, pp. 341–357, 2020.
- [18] J. A. S. JADHAV and J. DESHMUKH, “ADVANCING MACHINE LEARNING IN COVID-19 DIAGNOSTICS: SYMPTOM-BASED CLASSIFICATION AND ENSEMBLE TECHNIQUES,” *South East Eur J Public Health*, pp. 3044–3061, 2025.