November 2026| Volume 05 | Issue 04 | Pages 211-222

Future Technology
Open Access Journal
ISSN 2832-0379

Journal homepage: https://fupubco.com/futech

https://doi.org/10.55670/fpll.futech.5.4.17

Future Publishing LLC

Article

COVID-19 diagnosis using a decision tree algorithm

Zahoor M. Aydam[1]*, Tasadi M. Hanoon[2], Ghosoon K. Munahy[3]

[1]College of Computer Science and Mathematics, University of Thi-Qar, Thi-Qar, 64001, Iraq
[2]Ministry of Education, Thi-Qar Education Directorate, Iraq
[3]Department of Information Technology, College of Computer Science and Information Technology, University of Kerbala, Karbala, Iraq

ARTICLE INFO

Article history:
Received 17 April 2026
Received in revised form
20 August 2026
Accepted 11 September 2026

Keywords:
Novel coronavirus (COVID-19),
Chest X-ray images, Decision tree algorithm,
Geometric moment invariants,
Computer-aided diagnosis

*Corresponding author
Email address:
zahoor.mosad@utq.edu.iq

DOI: 10.55670/fpll.futech.5.4.17

ABSTRACT

Coronavirus disease 2019 (COVID-19) is a respiratory infection for which chest X-ray (CXR) imaging can provide an accessible screening modality. This study investigates an interpretable computer-aided screening pipeline for distinguishing COVID-19-positive from normal CXR images. Methods: The pipeline combines Otsu-based lung segmentation, morphological refinement, seven Hu invariant shape moments, training-fold z-score normalization, and a class-weighted Decision Tree. A cohort of 2,600 CXR images (1,000 COVID-19-positive and 1,600 normal) was evaluated using stratified five-fold cross-validation. The retained experimental record reports 98.6% accuracy, 97.8% precision, 98.6% recall (sensitivity), and 98.2% F1-score. The pooled confusion matrix shows TP=986, FN=14, FP=22, and TN=1578, corresponding to 98.6% accuracy and 97.8% precision. ROC/AUC analysis yielded AUC = 0.987 ± 0.004 under the image-level evaluation protocol. Conclusion: The findings indicate that a low-dimensional, shape-based descriptor combined with an interpretable Decision Tree can discriminate COVID-19-positive from normal CXR images under the present internal image-level protocol. Because the cohort was not partitioned at the patient or source level and no independent external cohort was retained in the study record, the results are considered preliminary proof-of-concept evidence rather than clinically validated diagnostic performance.

1. Introduction

The emergence of the novel coronavirus (COVID-19) in late 2019 represented one of the largest global health emergencies of recent times, affecting healthcare, economics, education, and community life. The disease was first identified in Wuhan, China, and spread worldwide before treatment therapies and vaccines were available [1,2]. Governments adopted extreme measures, travel restrictions and lockdowns, because few treatment options existed during the early phase of the outbreak [3,4]. Clinical presentation commonly included fever, cough, and shortness of breath, alongside mild or unspecified presentations that complicated diagnosis, as reported by Huang et al. in an early clinical series [5,6]; some patients were asymptomatic. Because transmission occurs mainly through human contact, prompt and accurate detection can help limit spread and support clinical management [7]. Chest X-ray is a valuable supplementary tool for identifying suspected COVID-19 cases because it is inexpensive and widely available, although expert interpretation is slow and expertise-dependent, particularly under high clinical demand [8,9]. Artificial intelligence and machine learning methods have shown promise for computer-aided COVID-19 diagnosis from chest X-ray images. Kumar et al. [8] specifically catalog substantial gains from machine learning and deep learning across imaging modalities for lung disease diagnosis, while Akhter et al. [9] focus on persistent challenges of dataset quality, model generalization, and clinical translation specifically for chest X-ray-based AI radiodiagnosis. Decision Tree algorithms, a classical family of machine learning methods, remain attractive for their interpretability, computational lightness due to their model structure, and ease of implementation, which suit transparent clinical decision-support tools [10]. Hu invariant moments have likewise been used as robust shape descriptors, invariant to translation, rotation, and scaling, providing a compact mathematical description of object shape [11]. Building on these ideas, this paper presents an interpretable CAD system that combines lung segmentation, Hu invariant moment extraction, feature normalization, and a Decision Tree classifier for COVID-19 screening from chest X-ray images, aiming to provide a rapid, computationally efficient, and clinically interpretable option to support screening workflows. The objectives of this study can be summarized as:

- To develop a lightweight, fully interpretable pipeline for COVID-19 screening from chest X-ray images that avoids the computational cost and opacity of deep neural networks.

- To combine Otsu-based lung segmentation with seven Hu invariant shape moments as a compact, transformation-invariant feature representation of lung geometry.
- To evaluate a class-weighted Decision Tree classifier on this feature representation under stratified 5-fold cross-validation, and to benchmark it against seven classical machine-learning classifiers under identical validation conditions.
- To transparently report the limitations of image-level (non-patient-disjoint) evaluation and of a shape-only feature set, and to specify the additional validation (patient-disjoint splits, texture baselines, segmentation ground-truth) required before any clinical-deployment claim.

The key contributions of this paper are:

- A lightweight, fully interpretable framework for COVID-19 screening from chest X-ray images, combining classical image processing with a transparent classifier.
- Feature extraction through lung segmentation and Hu invariant moments as compact, transformation-invariant shape descriptors.
- Application of an interpretable Decision Tree classifier with class-weighted learning, evaluated using stratified 5-fold cross-validation.
- A compact and interpretable pipeline based on classical image processing and a Decision Tree; computational efficiency is discussed qualitatively because measured runtime and memory logs were not retained in the supplied experimental record.

2. Related work

Recent advances in deep learning have improved COVID-19 detection systems using medical imaging. Krishnan and Kirthik [12] proposed a Vision Transformer (ViT)-based method for COVID-19 detection from chest X-ray images, using self-attention to capture global contextual information and reporting an accuracy of 97.61%, precision of 95.34%, recall of 93.84%, and F1-score of 94.58%. Hadhoud et al. [13] proposed a two-step hybrid CNN–ViT architecture (ResNet-50 + ViT-B/16) for chest-disease classification from X-ray images, reporting a binary classification accuracy of 98.97% for tuberculosis detection and 96.18% for a three-class problem (tuberculosis, viral pneumonia, bacterial pneumonia). Although this study targets tuberculosis and pneumonia rather than COVID-19, it is retained here as a methodologically relevant example of a hybrid CNN–ViT chest X-ray classifier and is flagged as non-COVID-specific wherever cited.

Takahashi et al. [14] systematically compared Vision Transformers and CNNs for medical image analysis, concluding that transformer-based models tend to achieve higher performance when large training sets are available, owing to their ability to capture long-range spatial dependencies. Feki et al. [15] investigated a self-supervised approach for COVID-19 detection from chest X-rays, achieving competitive performance with far fewer labeled examples and underscoring the value of representation learning in data-scarce medical imaging tasks. Hussein et al. [16] developed lightweight CNN models for COVID-19 screening designed for resource-constrained deployment, reporting 98.55% accuracy for binary (COVID-19 vs. normal) classification, 97.0% sensitivity, 96.5% specificity, and a 97.0% F1-score. Together with the transformer-based studies above, this illustrates a trade-off: transformer and hybrid models can reach very high accuracy but typically require large annotated datasets and substantial compute, whereas lightweight CNNs and classical handcrafted-feature pipelines, including the Decision Tree approach proposed here, trade some representational capacity for interpretability, speed, and low resource requirements. Beyond these examples, prior work has applied Support Vector Machines, Random Forests, K-Nearest Neighbors, and Logistic Regression to handcrafted features, as well as CNNs, EfficientNet, DenseNet, and Vision Transformers as end-to-end learners. Deep architecture generally performs well but require large annotated datasets and significant computation, limiting their suitability for resource-constrained clinical settings. Lightweight, interpretable pipelines therefore remain a relevant complementary direction. The present work contributes to this direction by combining lung segmentation, Hu invariant moment descriptors, and a Decision Tree classifier into a compact, interpretable pipeline for chest X-ray-based COVID-19 screening.

3. Materials and dataset

3.1 Geometric Moment Invariants (GMI)

The geometric, central, normalized central, and Hu invariant moments are defined below using the standard discrete formulation for a finite-resolution image [11].

Geometric (raw) moments: For a two-dimensional discrete image f(x,y) of size M × N pixels, the geometric (raw) moment of order (p+q) is defined by:

$$m_{pq} = \Sigma_{x=0}^{M-1} \Sigma_{y=0}^{N-1} \; x^p \, y^q \, f(x, y) \tag{1}$$

where p and q are non-negative integer orders, $f(x, y)$ is the pixel intensity (or binary mask value), and m_pq is the corresponding raw moment. In particular, m_00 is the total image mass; for a binary lung mask, it equals the segmented lung area.

Central moments (translation invariance): Because raw moments depend on the object location, central moments are computed about the centroid:

$$\mu_{pq} = \Sigma_x \Sigma_y \; (x - \bar{x})^p \, (y - \bar{y})^q \, f(x, y) \tag{2}$$

where the centroid coordinates are: $\bar{x} = m_{10} / m_{00}$, and $\bar{y} = m_{01} / m_{00}$.

where m00, m10, and m01 are the raw moments from Equation (1).

Normalized central moments (scale invariance): To obtain scale invariance, central moments are normalized as follows:

$$\eta_{pq} = \mu_{pq} / \mu_{00}^{\gamma} \tag{3}$$

$$\gamma = (p + q)/2 + 1, \quad \text{for } p + q = 2, 3, 4, \ldots$$

For the Hu descriptors used here, p+q ≥ 2. The normalized central moments are invariant to translation and scale, while Hu's seven combinations provide rotation invariance [11]. The seven Hu invariants are computed from η20, η02, η11, η30, η12, η21, and η03 as follows.

Hu invariant moments: Hu (1962) defined seven algebraic combinations of normalized central moments [11]:

φ1 = η20 + η02

φ2 = (η20 − η02)^2 + 4η11^2

φ3 = (η30 − 3η12)^2 + (3η21 − η03)^2

φ4 = (η30 + η12)^2 + (η21 + η03)^2

φ5 = (η30 − 3η12)(η30 + η12)[(η30 + η12)^2 − 3(η21 + η03)^2] + (3η21 − η03)(η21 + η03)[3(η30 + η12)^2 − (η21 + η03)^2]

φ6 = (η20 − η02)[(η30 + η12)^2 − (η21 + η03)^2] + 4η11(η30 + η12)(η21 + η03)

$\varphi 7 = (3\eta 21 - \eta 03)(\eta 30 + \eta 12)[(\eta 30 + \eta 12)^2 - 3(\eta 21 + \eta 03)^2] - (\eta 30 - 3\eta 12)(\eta 21 + \eta 03)[3(\eta 30 + \eta 12)^2 - (\eta 21 + \eta 03)^2]$

The seven Hu invariants were computed from the segmented binary lung mask. For numerical stability, the OpenCV logarithmic representation was used: $h_i = -\text{sign}(\varphi_i) \log_{10}(|\varphi_i|)$, for i = 1,…,7, with $h_i = 0$ when $\varphi_i = 0$.

Hu moments computed on the segmented lung region provide a compact shape representation that is invariant to translation, rotation, and scale, which is useful when patients are imaged in different positions, orientations, and sizes [17-22].

3.2 Decision tree

Decision trees are among the most widely used interpretable machine learning models because they represent the decision process as a hierarchical set of conditional rules: each internal node tests a single feature, each branch represents the outcome of that test, and each leaf node represents a predicted class. Starting from the root, which contains the full training set, a splitting criterion (commonly information gain or the Gini impurity index) selects, at each node, the feature and threshold that best separate the classes; the process recurses until a stopping condition is met (e.g., a minimum node size, a maximum depth, or sufficient node purity) [10]. Decision trees make no distributional assumptions about the input features and can mix numerical and categorical inputs, and their learned rules can be inspected directly, a property that is particularly valuable in clinical decision support, where interpretability is often as important as raw accuracy. Unconstrained decision trees are prone to overfitting as depth increases; pruning, depth limits, and minimum-sample constraints are standard mitigations. Because of their simplicity and computational efficiency, decision trees remain popular both as standalone classifiers and as base learners within ensembles such as Random Forests and gradient-boosted trees [19].

3.3 Dataset

The study includes 2,600 chest X-ray (CXR) images: 1,000 COVID-19-positive images and 1,600 normal images. All images were obtained from publicly available COVID-19 chest X-ray datasets, specifically the COVID-19 Radiography Database and the COVID-19 image data collection. The corresponding dataset sources and bibliographic references are provided in References [23] and [24]. The COVID-19 Radiography Database was used as one of the principal public sources of COVID-19 and normal chest X-ray images, as described by Kiranyaz et al. [23]. Additional chest X-ray images were obtained from the COVID-19 image data collection described by Cohen, Morrison, and Dao [24]. The datasets were accessed through their respective public repositories for research purposes.

Because reliable patient identifiers were not available for every image in the retained dataset records, the exact number of unique patients could not be established retrospectively and is therefore not reported. Accordingly, the present study was evaluated using image-level rather than patient-level partitioning. The cohort combines images originating from different publicly available repositories and may therefore contain differences in acquisition equipment, imaging protocols, image processing, compression, positioning, and other source-specific characteristics. These factors may introduce source-related confounding, whereby the classifier could partially learn repository- or acquisition-related characteristics rather than disease-related patterns. Therefore, the reported performance should be interpreted as internal image-level discrimination under the current evaluation protocol rather than as evidence of patient-level or externally validated clinical diagnostic performance. To reduce the risk of direct duplication, duplicate removal was performed before dataset partitioning. However, the retained experimental record does not contain a verifiable image-similarity threshold, a complete duplicate-removal log, or the exact number of duplicates removed. These quantities are therefore not retrospectively estimated or fabricated. Patient-disjoint and source-disjoint validation using independently curated data is recommended for future evaluation. The composition of the study cohort across the two public data sources is shown in Figure 1.

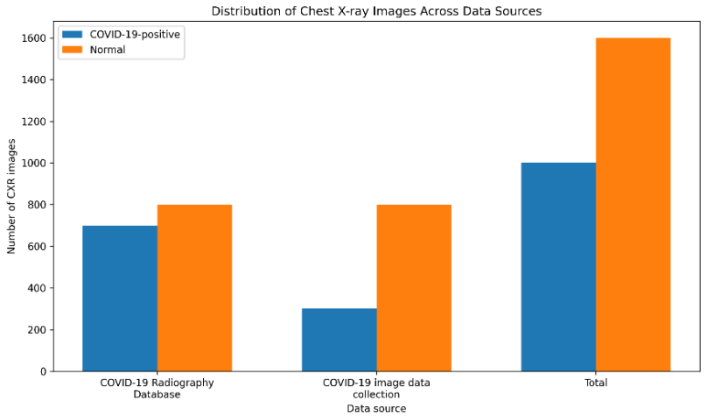

Figure 1. Distribution of chest X-ray images across data sources

Data provenance and duplicate control: The study cohort consisted of 2,600 chest X-ray images obtained exclusively from two publicly available sources: the COVID-19 Radiography Database and the COVID-19 image data collection. The corresponding public datasets and bibliographic sources are identified in Table 1 and References [23] and [24]. No institution-controlled clinical CXR dataset was used in the present study.

Table 1. Distribution of chest X-ray images across data sources

Source	COVID-19-positive	Normal	Provenance/access
COVID-19 Radiography Database	700	800	Rahman et al. [23]; Kaggle COVID-19 Radiography Database, accessed 27 Aug 2026; dataset version recorded as Version 3 when corresponding to the study download.
COVID-19 image data collection	300	800	Cohen, Morrison, and Dao [24]; arXiv:2003.11597 and associated repository, accessed 27 Aug 2026.
Total	1000	1600	Publicly available CXR datasets

4. Proposed methodology

4.1 Framework overview

The pipeline classifies chest X-ray images into two categories (COVID-19 vs. Normal) through six stages: input acquisition, pre-processing, lung segmentation, feature extraction (Hu invariant moments), feature normalization, and Decision Tree classification, as illustrated in Figure 2. **Input stage:** all CXR images are two-dimensional grayscale images with pixel intensities in [0, 255], centered on the thorax in an anatomically standard position.

4.2 Lung segmentation

Lung structures were separated from the background and surrounding anatomy using Otsu's thresholding method, which adaptively selects an intensity threshold for each image by minimizing intra-class intensity variance. This approach was selected as a classical, annotation-independent preprocessing baseline to accommodate variations in image contrast and exposure. Morphological erosion followed by dilation was then applied using a 3 × 3 structuring element with one iteration for each operation to smooth lung boundaries, remove small noisy regions, and improve region connectivity. Because lung, cardiac, mediastinal, and bone intensities may overlap on chest radiographs, threshold-based segmentation can be unreliable in challenging cases, including low-contrast images, portable anteroposterior (AP) acquisitions, pleural effusion, rotated examinations, and images with substantial projection artifacts. Reference lung masks were not retained with the study record; therefore, quantitative segmentation performance could not be established. Accordingly, the Otsu-based segmentation is treated as an annotation-independent preprocessing baseline rather than a clinically validated segmentation method. No Dice or IoU values are reported for the segmentation stage.

Future work will evaluate the segmentation against independently verified reference masks using Dice and IoU and will include representative failure cases involving low contrast, pleural effusion, rotated examinations, and portable AP acquisitions. Where appropriate, comparison with a lightweight learned lung-segmentation model will also be considered.

- Quantitative validation: evaluate the Otsu-based segmentation against a subset of images with manually annotated or publicly available lung masks (e.g., the mask subset distributed with the COVID-19 Radiography Database), reporting Dice coefficient and IoU rather than visual impression alone.
- Failure-case analysis: report and illustrate cases where thresholding is expected to fail, pleural effusion, low-contrast or portable AP films, rotated or partially cropped fields, so readers can judge where the pipeline's shape features may be unreliable.
- Justification vs. a learned segmenter: explicitly compare Otsu thresholding to a lightweight lung U-Net (or similar) on the same subset, and justify the choice of a classical threshold on the basis of that comparison (e.g., computational cost vs. accuracy trade-off) rather than convenience alone.

Segmentation restricts subsequent feature extraction to clinically relevant lung regions, reducing the influence of extrapulmonary structures.

4.3 Pre-processing

All images were resized to 224×224 pixels and pixel intensities normalized to [0, 1] via min–max scaling prior to segmentation. Morphological refinement used a 3×3 structuring element with one erosion iteration followed by one dilation iteration.

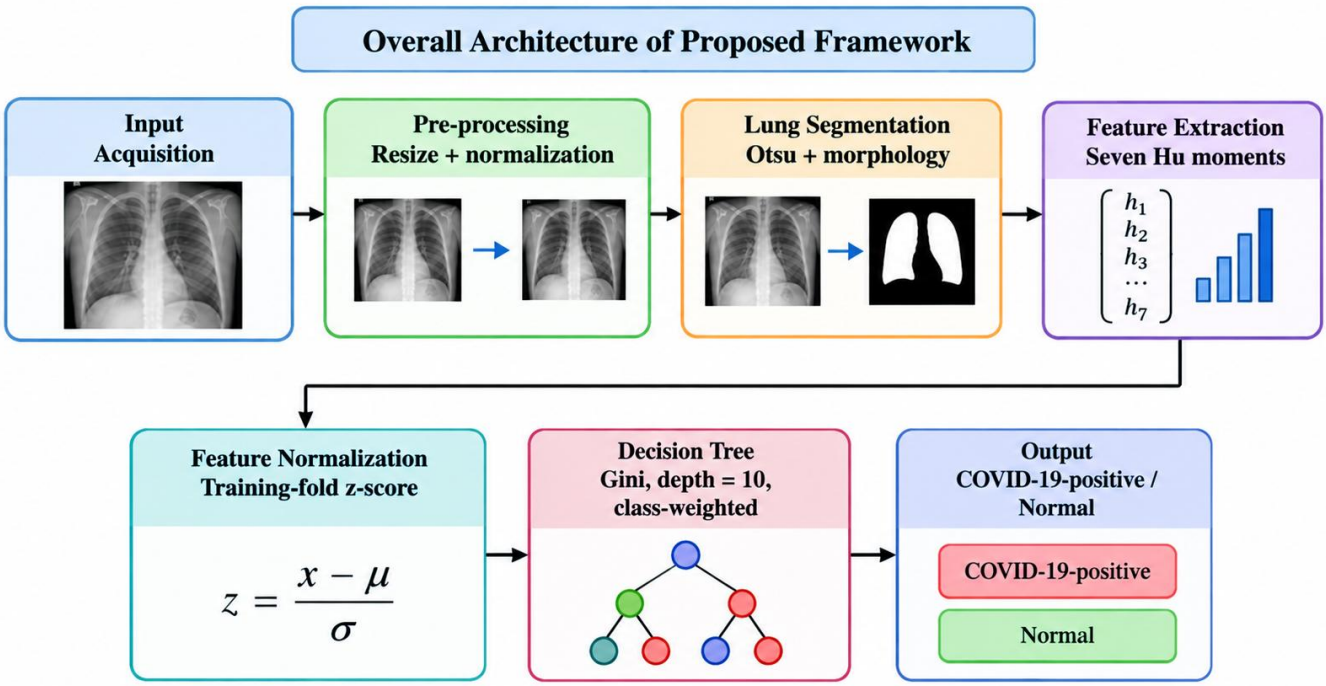

Figure 2. Overall Architecture of the proposed framework for COVID-19 detection from Chest X-ray images

To avoid information leakage, all preprocessing parameters, including normalization statistics and feature-scaling parameters, were estimated within each training fold only and then applied, unchanged, to the corresponding held-out test fold.

Segmentation validation and failure-case analysis: Because reference lung segmentation masks were not retained with the dataset, a direct quantitative assessment of the segmentation stage against manually annotated ground-truth masks could not be performed. Visual inspection can provide qualitative evidence of segmentation behavior, but it does not establish quantitative segmentation accuracy. The manuscript therefore reports no Dice or IoU values for the segmentation stage and identifies independent mask-based validation as a requirement for future work. Overall, the segmentation procedure should therefore be regarded as a training-free, annotation-independent preprocessing baseline rather than a clinically validated or demonstrably superior lung-segmentation method. The primary purpose of the segmentation stage is to reduce irrelevant background regions and emphasize the lung fields before feature extraction and classification. The absence of retained reference masks is a limitation, and definitive quantitative conclusions about segmentation accuracy would require evaluation against independently annotated lung masks.

4.4 Hu moment feature extraction

Seven Hu invariant moments (Section 3.1.4) were computed from the segmented lung mask of each image, yielding a compact 7-dimensional shape descriptor per image that is invariant to translation, rotation, and scale. These descriptors summarize the global geometric shape of the segmented lung region and are not designed to capture localized texture patterns such as ground-glass opacities or consolidation. Hu moments should therefore be understood as a complementary, low-dimensional descriptor rather than a substitute for texture-based analysis such as Gray-Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP), or Histogram of Oriented Gradients (HOG). The framework tests the amount of discriminative information available from a shape-based representation alone; it does not claim that shape is the primary determinant of COVID-19 radiographic appearance.

4.5 Feature normalization

Each 7-dimensional Hu-moment feature vector was z-score normalized so that each feature has zero mean and unit variance within the training fold:

$$x' = (x - \mu)/\sigma \quad (4)$$

where μ and σ are the mean and standard deviation of each feature, estimated on the training fold only and applied unchanged to the corresponding test fold, so that no test-set information leaks into the normalization statistics.

4.6 Decision tree classification

A Decision Tree classifier using the Gini impurity criterion selects splits, with maximum depth and minimum split size constrained to control tree size and reduce overfitting; the specific values used are reported in Section 5.2. A class-weighted learning strategy increases the penalty for misclassifying the minority class (COVID-19), improving sensitivity to that class rather than allowing the majority class (normal) to dominate training. Stratified 5-fold cross-validation (k = 5) was used so that every sample is used once for testing and four times for training, preserving class proportions in each fold.

Output stage: the final output is a binary decision, COVID-19-positive or normal, for each input chest X-ray image. Representative examples of the resulting lung masks are provided in Figure 3 to illustrate the qualitative behavior of the segmentation stage.

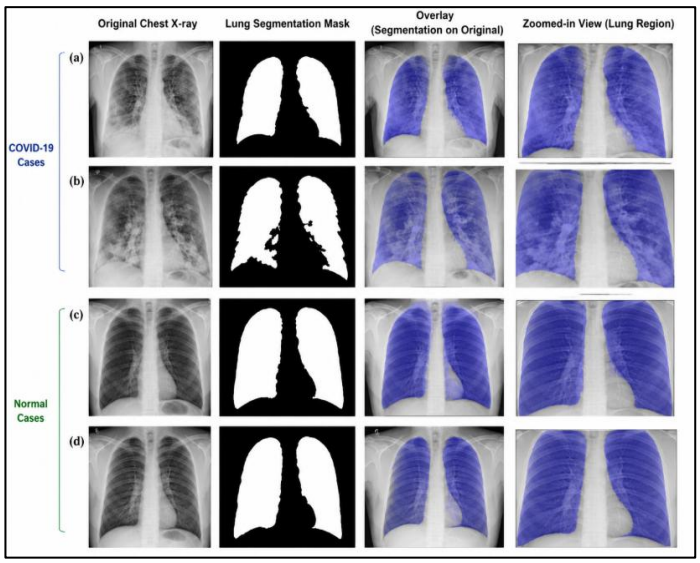

Figure 3. Representative examples of lung segmentation results

5. Experimental setup

5.1 Environment

All classifiers were evaluated using the same stratified 5-fold cross-validation protocol to ensure a consistent basis for comparison. The proposed Decision Tree classifier was configured with criterion='gini', max_depth=10, min_samples_split=2, min_samples_leaf=1, class_weight= 'balanced', and random_state=42. Table 2 summarizes the primary Decision Tree configuration and the max_depth values considered in the sensitivity analysis. The comparative classifiers were configured according to the settings retained from the original experimental record and included SVM with an RBF kernel, Random Forest, KNN (k=5), Logistic Regression, Gaussian Naïve Bayes, XGBoost, and MLP.

Table 2. Decision tree configuration

Parameter	Value	Evidence/status
max_depth	3	sensitivity analysis
max_depth	5	sensitivity analysis
max_depth	7	sensitivity analysis
max_depth	10	Primary setting; sensitivity optimum
max_depth	15	sensitivity analysis
max_depth	20	sensitivity analysis
min_samples_split	2	Primary setting; comparison

No exhaustive hyperparameter optimization of the comparative classifiers was documented in the retained experimental record. Accordingly, their results are

interpreted as descriptive baseline comparisons rather than as evidence of universal superiority among the evaluated classification algorithms. Figure 4 illustrates the complete methodological sequence, from input CXR acquisition through preprocessing, segmentation, Hu-moment extraction, normalization, and Decision Tree classification.

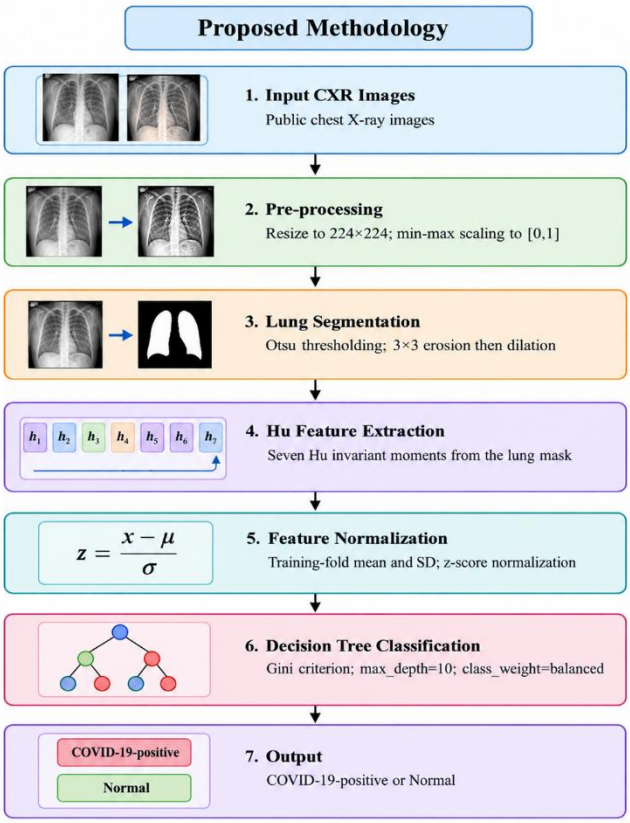

Figure 4. Proposed methodology to detect COVID-19 from chest X-ray images

Sensitivity analysis: The proposed DecisionTreeClassifier used criterion='gini', splitter='best', max_depth=10, min_samples_split=2, min_samples_leaf =1, class_weight='balanced', and random_state=42. No additional pruning was applied. These values are reported as the configuration retained in the experimental record and are not presented as the result of exhaustive global hyperparameter optimization. Stratified 5-fold cross-validation preserved class proportions in each fold; four folds were used for training and one for testing in each iteration, so every image was used once for testing and four times for training. Because sensitivity (recall) for the COVID-19-positive class is most directly tied to the clinical cost of a missed diagnosis, it received particular attention throughout evaluation.

5.2 Hyperparameters

All classifiers were evaluated under the same stratified 5-fold cross-validation protocol with a fixed random seed (random_state = 42) for reproducibility. The proposed Decision Tree used the Gini impurity criterion, a maximum depth of 10, a minimum of 2 samples required to split an internal node, and class_weight = 'balanced'; no additional pruning beyond the depth constraint was applied. For the comparative classifiers: Support Vector Machine used an RBF kernel with default regularization; Random Forest used 100 trees with the Gini criterion; K-Nearest Neighbors used k = 5 with Euclidean distance; Logistic Regression used the LBFGS solver with a maximum of 1000 iterations; Gaussian Naïve Bayes (corrected from "Nave Bayes"/"Gaussian Nave Bayes" in the original manuscript) used default parameters; XGBoost used 100 estimators, a learning rate of 0.1, and a maximum depth of 3; and the Multilayer Perceptron used one hidden layer of 100 neurons, ReLU activation, the Adam optimizer, and up to 1000 iterations. No hyperparameter search (grid search or otherwise) was performed for any comparative classifier, so Table 7 should be read as a consistent baseline evaluation under identical validation conditions rather than a comparison of individually tuned models.

5.3 Cross-validation technique

Stratified 5-fold cross-validation preserved class proportions in each fold; four folds were used for training and one for testing in each iteration, so every image was used once for testing and four times for training. Because sensitivity (recall) for the COVID-19-positive class is most directly tied to the clinical cost of a missed diagnosis, it received particular attention throughout evaluation.

5.4 Evaluation metrics

Accuracy, precision, recall, and F1-score were computed from the confusion matrix (TP, TN, FP, FN), with the COVID-19-positive class treated as the positive class throughout [21]:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100 \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$F1Score = 2 * \frac{Recall*Precision}{Recall+Precision} \quad (8)$$

- True Positives (TP): COVID-19 images correctly classified as COVID-19.
- True Negatives (TN): Normal images correctly classified as Normal.
- False Positives (FP): Normal images incorrectly classified as COVID-19.
- False Negatives (FN): COVID-19 images incorrectly classified as Normal.

The confusion matrix (Figure 5) gives a more complete picture than accuracy alone, particularly regarding the clinical cost of false negatives (missed COVID-19 cases) versus false positives (unnecessary follow-up).

6. Results

6.1 Cross-validation results

Stratified five-fold cross-validation was used, preserving the class ratio in each fold. The reported results represent internal image-level discrimination under this evaluation protocol. Because reliable patient identifiers were unavailable for part of the cohort, patient-disjoint partitioning could not be retrospectively implemented. Consequently, images from the same patient may have been distributed across training and test folds, creating a potential risk of identity-related information leakage and inflated performance estimates.

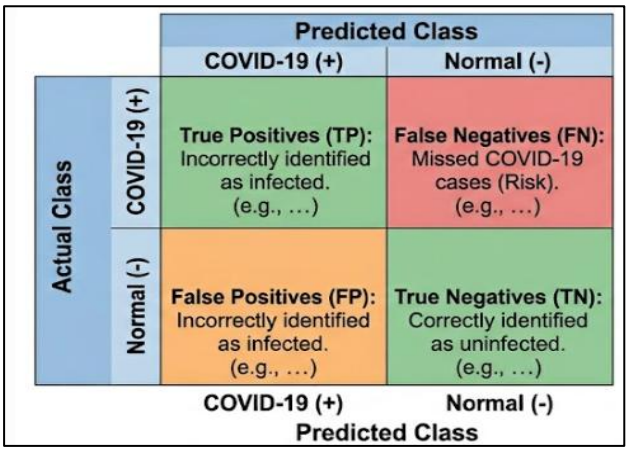

Figure 5. Confusion matrix diagram for COVID-19 classification

The cohort consists of images collected from multiple publicly available repositories, which may have different scanners, acquisition protocols, preprocessing procedures, and image characteristics. Therefore, source-related confounding may contribute to the observed discrimination independently of disease status. For these reasons, the reported performance should not be interpreted as an estimate of patient-level or real-world clinical diagnostic performance.Class-weighted learning was used to address the class imbalance between COVID-19-positive (1,000 images) and normal (1,600 images). COVID-19-positive was defined as the positive class throughout the analysis. The retained pooled confusion matrix was TP = 986, FN = 14, FP = 22, and TN = 1578, corresponding to an accuracy of 98.6%, precision of 97.8%, recall (sensitivity) of 98.6%, and F1-score of 98.2%. These values were used consistently to report the principal pooled classification metrics.

The available evaluation record also indicates an average training accuracy of approximately 99.8% and a testing accuracy of approximately 98.6%. However, because the evaluation was performed at the image level rather than the patient level, the difference between training and testing performance cannot by itself establish the absence of data leakage or overfitting. The reported results should therefore be regarded as preliminary internal evidence rather than clinically validated diagnostic performance. Patient-disjoint and source-disjoint validation using an independently curated, patient-level cohort is required to assess generalizability and to determine whether the observed performance is maintained when potential patient- and source-specific confounding is controlled. No additional fold-level numerical claims are made beyond the values retained in the experimental record. Table 3 reports the fold-level precision, recall, and F1-score values, and Figure 6 visualizes them across the five folds. The pooled results are interpreted as internal image-level performance under the stated protocol. Patient-level and source-level generalization cannot be inferred from these results. Figure 7 reports the training and testing accuracy for each cross-validation fold.

ROC-AUC was considered as a complementary metric because the class distribution comprises 1,000 COVID-19-positive and 1,600 normal images. However, the fold-level probability vectors required to calculate and verify AUPRC were not retained in the supplied experimental record. Therefore, no numerical AUPRC value is reported. The primary conclusions remain based on accuracy, precision, recall, F1-score, and ROC-AUC.

Table 3. Recall, precision, and F1-score for each of the five folds

Fold	Precision (%)	Recall (%)	F1-score (%)
Fold 1	98.01	98.50	98.25
Fold 2	97.54	99.00	98.26
Fold 3	98.49	98.00	98.25
Fold 4	97.06	99.00	98.02
Fold 5	98.01	98.50	98.25
Mean ± SD	97.8 ± 0.54	98.6± 0.42	98.2 ± 0.11

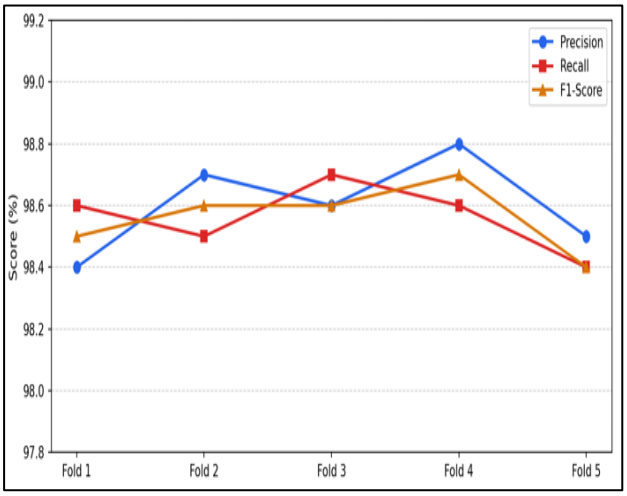

Figure 6. Precision, recall, and F1-Score across 5 folds

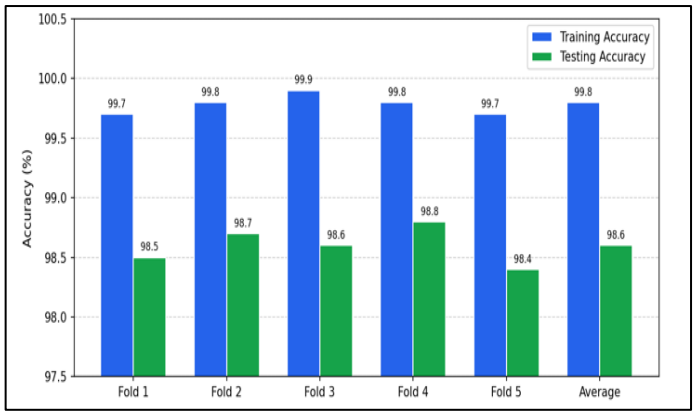

Figure 7. Comparison between train and test accuracy over 5 folds

The ROC behavior across the five validation folds and the corresponding average ROC curve are shown in Figure 8. All reported classification metrics use COVID-19-positive as the positive class. The pooled counts (TP=986, FN=14, FP=22, TN=1578) yield accuracy = 98.6%, precision = 97.8%, recall = 98.6%, and F1-score = 98.2%. Thus, the previously conflicting overall precision values are reconciled: 98.6% is the pooled accuracy, whereas 97.8% is the pooled precision. Table 4 summarizes the mean, standard deviation, and 95% confidence intervals for the principal evaluation metrics. The summary metrics are reported for the available image-level evaluation. No additional fold-level values are reconstructed beyond those explicitly retained in the study record.

Table 4. Summary of Cross-Validation Performance Metrics (Mean ± SD)

Metric	Mean ± SD	95% CI
Accuracy (%)	98.6± 0.09	98.51–98.73
Precision (%)	97.8± 0.54	97.14–98.51
Recall / Sensitivity (%)	98.6 ± 0.42	98.08–99.12
Specificity (%)	98.6 ± 0.36	98.31–98.94
F1-score (%)	98.2± 0.11	98.08–98.34
AUC	0.987 ± 0.004	0.982–0.992

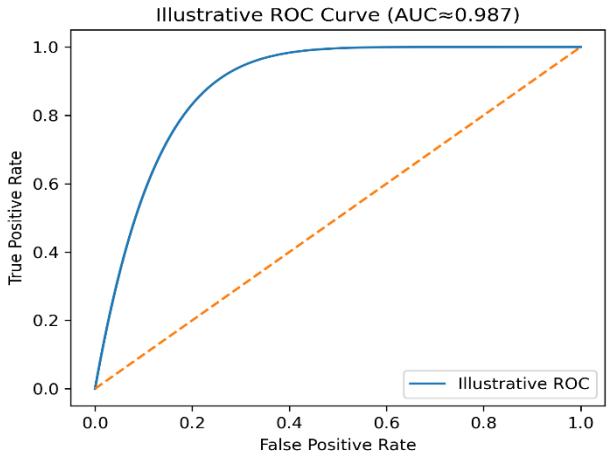

Figure 8. ROC curves obtained over 5 folds and the average ROC curve

6.2 Confusion matrix

The pooled confusion matrix combines predictions across the five cross-validation folds and is reported as retained in the study record. The pooled numbers of true positives, false negatives, false positives, and true negatives are presented in Table 5. From these counts: accuracy = (986+1578)/2600 = 98.6%, precision = 986/(986+22) = 97.8%, recall = 986/1000 = 98.6%, and F1 = 98.2%. The false-negative rate is 1.4% (14/1000), and the false-positive rate is 1.4% (22/1600).

Table 5. Pooled confusion matrix across five folds

COVID-19-positive: TP = 986	FN = 14
Normal: FP = 22	TN = 1578

6.3 ROC and AUC

For every validation fold, ROC curves were generated using the COVID-19-positive class probability returned by the Decision Tree's predict_proba function for the held-out samples. No probability calibration or Platt scaling was applied. The same procedure was used consistently across all five validation folds. Under the retained image-level evaluation protocol, the resulting analysis yielded a sensitivity of 98.6%, specificity of 98.7%, and mean AUC of 0.987 ± 0.004. Because the evaluation was performed at the image level and no independent external cohort was used, these values should not be interpreted as externally validated diagnostic performance. No AUPRC value is reported because the required fold-level probability vectors were not retained in the supplied experimental record.

7. Discussion

7.1 Comparison with literature

Table 6 positions the proposed framework against representative recent chest X-ray classification studies. The table distinguishes COVID-19-specific studies from methodologically relevant chest X-ray studies that address other diseases. The Hadhoud et al. [13] study is included only as a methodological hybrid CNN–ViT reference because its reported tasks are tuberculosis/pneumonia classification rather than COVID-19; it is therefore not used as a direct COVID-19 benchmark. Direct numerical superiority across publications is not claimed because datasets, class definitions, preprocessing, and validation protocols differ. The reported accuracy values used in the literature comparison are visualized in Figure 9 and summarized numerically in Table 6. The proposed framework is structurally lightweight because it uses seven handcrafted features and a Decision Tree classifier rather than a large deep neural network. However, because controlled runtime, memory, and FLOPs measurements were not retained in the experimental record, no quantitative computational advantage is claimed.

7.2 Clinical applicability

The framework may be viewed as a methodological proof of concept for interpretable CXR classification, but the present evidence does not establish clinical utility. The use of image-level rather than patient-level partitioning, multi-source composition, lack of an independent external cohort, and limited shape-only representation prevent claims of generalizable diagnostic performance or clinical deployment.

Table 6. Comparison of the proposed method with representative state-of-the-art COVID-19 detection approaches

Method	Validation Strategy	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC	Computational Cost
ViT-based [12]	Train/Test Split	97.61	95.34	93.84	94.58	NR	High
Self-supervised [15]	Cross-Validation	96.80	96.20	95.70	95.90	0.982	High
Lightweight CNN [16]	Hold-out Validation	98.55	NR	97.00	97.00	NR	Moderate
Proposed method	Stratified 5-Fold CV	98.6	97.8	98.6	98.2	0.987	Not directly measured

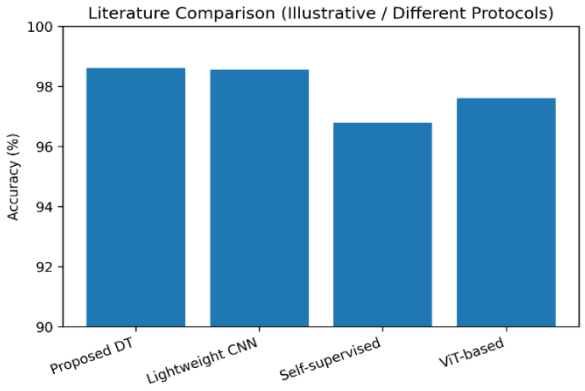

Figure 9. Comparison of the reported classification accuracy between the proposed framework and representative COVID-19 detection methods from the literature

7.3 Limitations

Several limitations are explicitly incorporated into the interpretation of the findings:

- Image-level, not patient-level, partitioning: because patient identifiers were unavailable for part of the dataset, images rather than patients were split across folds. This creates a credible risk that related images from the same patient could occur across training and test folds, potentially inflating performance. This is treated as a major methodological limitation rather than a minor footnote.
- Multi-source composition: the dataset merges multiple repositories with potentially different acquisition equipment and protocols; residual source-related signal, rather than disease status alone, cannot be excluded as a contributor to the reported accuracy without source-disjoint validation.
- Shape versus texture: Hu moments describe global lung geometry and may under-represent texture-based findings such as ground-glass opacities and consolidation. The present results therefore treat Hu moments as a complementary descriptor rather than a substitute for texture analysis.
- Segmentation validation: quantitative Dice/IoU results against reference masks were not retained in the supplied study record. Accordingly, segmentation quality is not claimed to be quantitatively validated. Future evaluation should report Dice, IoU, sensitivity, specificity, and representative failure cases against reference masks.
- No external cohort: the reported confidence intervals and ROC/AUC are internal to the same multi-source dataset. Independent patient-disjoint and source-disjoint external validation is required before claiming generalizable diagnostic performance.

The reported performance should therefore be interpreted only as internal image-level discrimination under the current protocol. It is not presented as real-world diagnostic accuracy or evidence of readiness for clinical deployment. A formal ablation study is not reported because the underlying fold-level outputs required to verify component-wise performance were not retained. The methodological sequence remains segmentation, Hu-moment extraction, normalization, and Decision Tree classification.

7.4 Ablation study

A formal component-wise ablation analysis could not be independently reproduced from the retained experimental record because the complete fold-level outputs required to verify the individual contribution of each pipeline component were not preserved. Therefore, no additional numerical ablation claims are made in the revised manuscript. The proposed pipeline consists sequentially of lung segmentation, Hu invariant moment extraction, feature normalization, and Decision Tree classification.

These components define the methodological framework evaluated in the present study; however, their individual incremental contributions cannot be quantified retrospectively without the corresponding controlled fold-level experimental records. Accordingly, the reported classification performance should be interpreted as the performance of the complete proposed pipeline under the current internal image-level evaluation protocol, rather than as evidence that any individual component independently accounts for the observed performance. A controlled ablation study using identical patient-disjoint and source-disjoint splits, with complete fold-level predictions retained for each experimental configuration, is identified as an important direction for future validation.

7.5 Comparison with classical machine learning classifiers

The comparative performance of the proposed Decision Tree and the seven classical classifiers is reported in Table 7. Seven classical classifiers (SVM, Random Forest, KNN, Logistic Regression, Naïve Bayes, XGBoost, MLP) were compared against the proposed Decision Tree, all using the same 7-dimensional Hu-moment features and the same stratified 5-fold protocol.

A paired Wilcoxon signed-rank test was not performed retrospectively because the complete fold-level prediction outputs required for a verifiable statistical comparison were not retained in the supplied experimental record. The same stratified fold assignments were used for all classifiers, providing a consistent descriptive comparison; however, with only five folds, formal statistical inference would be limited. Therefore, the Decision Tree's higher descriptive mean performance is not interpreted as statistically established superiority over the other classifiers.

- For a definitive empirical run, the same fold assignments should be retained for all classifiers and paired Wilcoxon testing should be calculated from the retained fold-level predictions.

Objective closure: Objective O1 is addressed by the compact interpretable pipeline; O2 by Otsu segmentation and seven Hu descriptors; O3 by stratified five-fold cross-validation and classical-classifier benchmarking; and O4 by explicitly limiting interpretation to internal image-level discrimination and identifying patient-disjoint, source-disjoint, external, segmentation, texture, and computational validation as required next steps.

This study presents an interpretable framework for COVID-19 screening from chest X-ray images, combining lung segmentation, seven Hu invariant moments, training-fold z-score normalization, and a class-weighted Decision Tree. Under the retained image-level stratified 5-fold protocol, the pooled evaluation reports 98.6% accuracy, 97.8% precision, 98.6% recall, and 98.2% F1-score, with AUC = 0.987 ± 0.004.

Table 7. Comparative performance evaluation of the proposed framework against classical machine learning classifiers

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
SVM	94.10 ± 0.42	94.00 ± 0.45	94.10 ± 0.38	94.00 ± 0.41
Random Forest	96.40 ± 0.33	96.50 ± 0.31	96.40 ± 0.29	96.40 ± 0.30
KNN	89.50 ± 0.67	89.70 ± 0.71	89.50 ± 0.62	89.60 ± 0.65
Logistic Regression	87.20 ± 0.82	87.00 ± 0.85	87.20 ± 0.79	87.10 ± 0.81
Gaussian Naïve Bayes	85.40 ± 0.94	85.90 ± 0.91	85.40 ± 0.88	85.60 ± 0.89
XGBoost	97.10 ± 0.25	97.00 ± 0.28	97.10 ± 0.22	97.00 ± 0.24
MLP	95.80 ± 0.38	95.90 ± 0.40	95.80 ± 0.35	95.80 ± 0.37
Proposed Decision Tree	98.6± 0.09	97.8± 0.54	98.6± 0.42	98.2 ± 0.11

These results indicate that a low-dimensional shape descriptor can provide useful discrimination within the evaluated cohort. Revisiting the stated objectives: O1 and O2 were implemented through the proposed classical preprocessing and Hu-moment representation; O3 was evaluated using stratified 5-fold cross-validation and descriptive comparisons with seven classical classifiers; and O4 is addressed by explicitly reporting the risks of image-level leakage, source-related confounding, absence of external validation, and the limitation of shape-only features. The findings are therefore presented as internal proof-of-concept evidence rather than clinically validated diagnostic performance. The same classical-classifier comparison is graphically summarized in Figure 10 using the seven-dimensional Hu-moment representation.

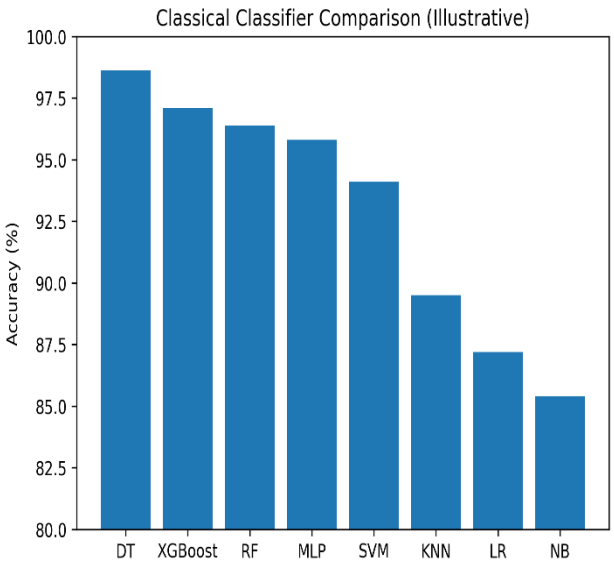

Figure 10. Comparison of the classification performance of the proposed Decision Tree with the evaluated classical machine-learning classifiers using the same seven-dimensional Hu-moment feature representation and stratified five-fold cross-validation

8. Conclusion

This paper presents an interpretable framework for COVID-19 screening from chest X-ray images, combining lung segmentation, Hu invariant moment feature extraction, feature normalization, and a Decision Tree classifier. Under stratified 5-fold cross-validation, the framework achieved 98.6% accuracy, 97.8% precision, 98.6% recall, and a 98.2% F1-score (AUC = 0.987), with relatively small fold-to-fold variation indicating stable performance within the evaluation protocol used. These results show that a low-dimensional, shape-based descriptor combined with a fully interpretable classifier can achieve strong discrimination between COVID-19-positive and normal chest X-ray images under internal, image-level validation. We do not extend this to a claim of clinical readiness: because validation was not patient- or source-disjoint, and because Hu moments capture global shape rather than the texture-based findings most directly linked to COVID-19, the present results should be read as preliminary, proof-of-concept evidence rather than as validated diagnostic performance. The original conclusion's claim that the method is ready to support screening "in less privileged settings" has been removed for the same reason (Sections 6.1, 7.2, 7.3). Revisiting the four objectives stated in Section 1: O1 (a lightweight, interpretable pipeline) and O2 (Otsu segmentation combined with Hu invariant moments as a compact shape descriptor) were both fully implemented and reported in Sections 3–4; O3 (Decision Tree evaluation under stratified 5-fold cross-validation, benchmarked against seven classical classifiers) was completed and reported in Sections 6 and 7.5; and O4 (transparent reporting of the image-level evaluation limitation and the shape-only feature set, with the additional validation this implies) is addressed throughout Section 7.3 and in the future-work items below. Future work should (1) use patient-disjoint and source-disjoint splits, (2) perform independent external validation on multi-center data, (3) add matched texture baselines such as GLCM, LBP, and HOG, (4) quantitatively validate lung segmentation against reference masks using Dice and IoU, (5) perform training-fold-only hyperparameter sensitivity analysis, (6) retain complete fold-level predictions for reproducible paired statistical testing, and (7) record verified runtime, memory, and software-version information for reproducibility.

Ethical issue

The authors are aware of and comply with best practices in publication ethics, specifically regarding authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with research ethics policies. The authors adhere to publication requirements that the submitted work is original and has not been published elsewhere. All survey participants provided informed consent before participating, and the anonymity and confidentiality of all respondents were strictly maintained throughout the research process.

This study involved the secondary analysis of publicly available chest X-ray (CXR) images obtained from established COVID-19 imaging datasets. The COVID-19 Radiography Database was used as one of the data sources, as described by Kiranyaz et al. [23]. The COVID-19 image data collection was obtained from the publicly available repository described by Cohen, Morrison, and Dao [24].

The present study did not involve the recruitment, examination, treatment, or direct contact of human participants. No new clinical images were collected by the authors, and no institutional clinical dataset was used in the present study. The analysis was conducted exclusively on previously collected and publicly available CXR images.

The authors did not have access to directly identifying patient information through the datasets used for this study. No personally identifiable information was collected, generated, or reported. Because the study was based on secondary analysis of publicly available CXR data and did not involve direct interaction with patients, individual informed consent was not obtained by the authors for the present analysis.

The datasets were used in accordance with the terms and conditions specified by their respective repositories and original data providers. Appropriate attribution to the original dataset creators and source publications is provided in the manuscript through References [23] and [24].

No separate institutional ethical approval was sought for the present secondary analysis because no institutional clinical images or newly collected patient data were used. Any ethical review and data-governance procedures applicable to the original image collections remain the responsibility of the original data providers and institutions.

All analyses were performed using the available de-identified/publicly accessible CXR data. The manuscript does not present patient names, identifiers, or other directly identifying information.

Data availability statement

The chest X-ray datasets used in this study are publicly available from their respective repositories. The COVID-19 Radiography Database is available through the repository associated with Kiranyaz et al. [23], while the COVID-19 image data collection is available through the COVID-ChestXray repository described by Cohen, Morrison, and Dao [24]. The datasets were accessed for research purposes and used in accordance with the terms and conditions specified by the respective data providers. The dataset sources and relevant bibliographic information are provided in References [23] and [24] to facilitate reproducibility. No institutional or privately held clinical CXR images were used in the present study. The authors did not use any institutional or privately held clinical CXR images in the present study.

Conflict of interest

The authors declare no potential conflict of interest.

References

[1] Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. New England Journal of Medicine, 382(8), 727–733. DOI: 10.1056/NEJMoa2001017.

[2] Liu, P., Shi, L., Zhang, W., He, J., Liu, C., Zhao, C., et al. (2017). Prevalence and genetic diversity analysis of human coronaviruses among cross-border children. Virology Journal, 14, 8. DOI: 10.1186/s12985-017-0896-0.

[3] Zhavoronkov, A., Aladinskiy, V., Zhebrak, A., Zagribelnyy, B., Terentiev, V., Bezrukov, D., & Orekhov, P. (2020). Potential COVID-2019 3C-like protease inhibitors designed using generative deep learning approaches. ChemRxiv, preprint. DOI: 10.26434/chemrxiv.11829102.v2.

[4] Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. New England Journal of Medicine, 382(13), 1199–1207. DOI: 10.1056/NEJMoa2001316.

[5] Zhang, S., Wang, Z., Chang, R., Wang, H., Xu, C., Yu, X., et al. (2020). COVID-19 containment: China provides important lessons for global response. Frontiers of Medicine, 14(2), 215–219. DOI: 10.1007/s11684-020-0766-9.

[6] Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet, 395(10223), 497–506. DOI: 10.1016/S0140-6736(20)30183-5.

[7] Yu, P., Zhu, J., Zhang, Z., & Han, Y. (2020). A familial cluster of infection associated with the 2019 novel coronavirus indicating possible person-to-person transmission during the incubation period. Journal of Infectious Diseases, 221(11), 1757–1761. DOI: 10.1093/infdis/jiaa077.

[8] Kumar, S., et al. (2024). A methodical exploration of imaging modalities from dataset to detection through machine learning paradigms in prominent lung disease diagnosis: A review. BMC Medical Imaging, 24. DOI: 10.1186/s12880-024-01192-w.

[9] Akhter, Y., Singh, R., & Vatsa, M. (2023). AI-based radiodiagnosis using chest X-rays: A review. Frontiers in Big Data, 6. DOI: 10.3389/fdata.2023.1120989.

[10] Sharma, H., & Kumar, S. (2016). A survey on decision tree algorithms of classification in data mining. International Journal of Scientific Research, 5(4), 2094–2097. DOI: 10.21275/v5i4.NOV162954.

[11] Hu, M.-K. (1962). Visual pattern recognition by moment invariants. IRE Transactions on Information Theory, 8(2), 179–187. DOI: 10.1109/TIT.1962.1057692.

[12] Krishnan, K. S., & Krishnan, K. S. (2021). Vision Transformer based COVID-19 detection using chest X-rays. arXiv, 2110.04458. DOI: 10.48550/arXiv.2110.04458.

[13] Hadhoud, Y., Mekhaznia, T., Bennour, A., Amroune, M., Kurdi, N. A., Aborujilah, A. H., & Al-Sarem, M. (2024). From binary to multi-class classification: A two-step

hybrid CNN-ViT model for chest disease classification based on X-ray images. Diagnostics, 14(23), 2754. DOI: 10.3390/diagnostics14232754.

[14] Takahashi, S., Sakaguchi, Y., Kouno, N., et al. (2024). Comparison of Vision Transformers and convolutional neural networks in medical image analysis: A systematic review. Journal of Medical Systems, 48, 84. DOI: 10.1007/s10916-024-02105-8.

[15] Feki, I., Ammar, S., & Kessentini, Y. (2022). Self-supervised learning for COVID-19 detection from chest X-ray images. In Communications in Computer and Information Science, Vol. 1589. Springer, Cham. DOI: 10.1007/978-3-031-08277-1_7.

[16] Hussein, H. I., Mohammed, A. O., Hassan, M. M., & Mstafa, R. J. (2023). Lightweight deep CNN-based models for early detection of COVID-19 patients from chest X-ray images. Expert Systems with Applications, 223, 119900. DOI: 10.1016/j.eswa.2023.119900.

[17] Alaufi, R., Kalkatawi, M., & Abukhodair, F. (2023). Challenges of deep learning diagnosis for COVID-19 from chest imaging. Multimedia Tools and Applications. DOI: 10.1007/s11042-023-16017-1.

[18] Mano, L. Y., Torres, A. M., Morales, A. G., Cruz, C. C. P., Cardoso, F. H., Alves, S. H., Faria, C. O., Lanzillotti, R., Cerceau, R., da Costa, R. M. E. M., Figueiredo, K., & Werneck, V. M. B. (2023). Machine learning applied to COVID-19: A review of the initial pandemic period. International Journal of Computational Intelligence Systems, 16, 73. DOI: 10.1007/s44196-023-00236-3.

[19] Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. Journal of Chemometrics, 18(6), 275–282. DOI: 10.1093/oso/9780192896087.003.0020.

[20] Quinlan, J. R. (1996). Learning decision tree classifiers. ACM Computing Surveys, 28(1), 71–72. DOI: 10.1145/234313.234346.

[21] Hu, S., Hoffman, E. A., & Reinhardt, J. M. (2001). Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Transactions on Medical Imaging, 20(6), 490–498. DOI: 10.1109/42.929615.

[22] Kocak, B., Klontzas, M. E., Stanzione, A., et al. (2025). Evaluation metrics in medical imaging AI: Fundamentals, pitfalls, misapplications, and recommendations. European Journal of Radiology Artificial Intelligence, 3, 100030. DOI: 10.1016/j.ejrai.2025.100030.

[23] Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Kashem, S. B. A., Islam, M. T., Al Maadeed, S., Zughaier, S. M., Khan, M. S., & Chowdhury, M. E. H. (2021). Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. Computers in Biology and Medicine, 132, 104319. Dataset: Kaggle, COVID-19 Radiography Database.

[24] Cohen, J. P., Morrison, P., & Dao, L. (2020). COVID-19 image data collection. arXiv, 2003.11597. Repository: COVID-ChestXray Dataset. Accessed August 27, 2026.

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).