# Translational application of a self-organized deep feature engineering pipeline for non-invasive pulmonary hypertension classification from routine chest radiographs

Intelligent Medicine xxx (xxxx) xxx

Contents lists available at ScienceDirect

Intelligent Medicine

journal homepage: www.elsevier.com/locate/imed

Research Article

# Translational application of a self-organized deep feature engineering pipeline for non-invasive pulmonary hypertension classification from routine chest radiographs

Tarık Kıvrak[1], Mehmet Ali Gelen[1], Ozge Salkin[2], Ozkan Karaca[3], Prabal Datta Barua[4], Sengul Dogan[5], Turker Tuncer[5], Ru-San Tan[6,7], Massimo Salvi[8,*], Vinitha Sree Subbhuraam[9], U.R. Acharya[10,11]

[1] *Department Cardiology, Firat University Hospital, Firat University, Elazig 23200, Turkey*
[2] *Department Cardiology, Adana Seyhan State Hospital, Adana 01430, Turkey*
[3] *Department Cardiology, Mersin City Training and Research Hospital, Mersin 33120, Turkey*
[4] *School of Business (Information System), University of Southern Queensland, Toowoomba QLD 4350, Australia*
[5] *Department of Digital Forensics Engineering, College of Technology, Firat University, Elazig 23119, Turkey*
[6] *Department of Cardiology, National Heart Centre Singapore, 169609, Singapore*
[7] *Duke-NUS Medical School, 169857, Singapore*
[8] *Department of Electronics and Telecommunications, Politecnico di Torino, Torino 10129, Italy*
[9] *The Digital Health Hub, Austin, Texas 78701, USA*
[10] *School of Mathematics, Physics and Computing, University of Southern Queensland, Toowoomba QLD 4350, Australia*
[11] *Centre for Health Research, University of Southern Queensland, Toowoomba QLD 4350, Australia*

ARTICLE INFO

*Keywords:*
Pulmonary hypertension
Chest X-Ray
Artificial intelligence
Machine learning
Deep learning
Explainable artificial intelligence
MobileNetV2
Support vector machine
k-nearest neighbors
Translational study

ABSTRACT

**Background** Pulmonary hypertension (PH) causes high mortality and poses diagnostic challenges. Current guidelines require invasive right heart catheterization (RHC) to confirm mean pulmonary artery pressure ≥25 mmHg. Delayed diagnosis impairs timely treatment. It is unknown whether standard chest X-rays can stratify PH severity. We aimed to develop and validate Exemplar MobileNet (ExMobileNet), an explainable artificial intelligence (AI) model that classifies PH into hemodynamic categories from routine chest X-ray images and thus supports non-invasive severity assessment.

**Methods** We collected 1,293 de-identified chest X-rays obtained from 2018 to 2023. The cohort comprised 135 patients with PH confirmed using RHC and 551 healthy controls. We defined seven multi-class tasks for key hemodynamic parameters (such as mean pulmonary artery pressure, pulmonary vascular resistance, and cardiac index). The ExMobileNet workflow consists of: (1) Feature extraction via MobileNetV2, (2) feature selection by neighborhood component analysis and chi-square feature selectors, (3) classification with k-nearest neighbors and support vector machines and (4) decision fusion by majority vote and greedy optimization.

**Results** Task-level accuracy ranged from 90.3% to 93.2%. Geometric mean scores ranged from 78.9% to 85.1%. Overall sensitivity and specificity were 88.5% and 91.3%, respectively. Mean accuracy across all tasks was 92.0% (±1.2%). Average inference time was 2.3 ± 0.4 second per image on CPU-only hardware.

**Conclusions** ExMobileNet achieved high agreement with RHC-based assessments using routine chest X-rays. This AI tool may enable earlier, non-invasive PH screening in clinical practice.

## 1. Introduction

Pulmonary hypertension (PH) is a chronic disease marked by high pressure in the lung circulation. It causes serious illness and high death rates [1,2]. Symptoms include breathlessness, chest pain, fatigue, and weakness [3]. Definitive diagnosis requires right heart catheterization (RHC) showing mean pulmonary artery pressure ≥25 mmHg [4,5]. RHC also guides treatment by classifying PH as pre-capillary, post-capillary, or mixed [6]. For instance, pulmonary arterial hypertension (PAH) [7] benefits from targeted drugs, whereas chronic thromboembolic PH (CTEPH) may be cured surgically [8,9].

Therefore, early non-invasive screening for PH is needed. Chest X-ray is widely available but underused for severity assessment [10–12]. Machine learning (ML) holds promise for automated detection and strat-

* Corresponding author: Massimo Salvi, Department of Electronics and Telecommunications, Politecnico di Torino, Torino 10129, Italy.
*E-mail address:* massimo.salvi@polito.it (Massimo Salvi).

https://doi.org/10.1016/j.imed.2025.06.003
Received 3 April 2025; Received in revised form 21 May 2025; Accepted 26 June 2025
Available online xxx

ification [13–17]. Prior ML studies encompass three main approaches: audio-based [18], electrocardiogram (ECG)-based [19], and imaging-based methods [20].

Audio-based models detect changes in heart sounds, with Wang et al [21] achieving 94.4% accuracy using wavelet-scattering networks on phonocardiograms. Ge et al [22] combined time-frequency domain analysis with a convolutional neural network (CNN) for detecting PH in patients with congenital heart disease. ECG-based models use deep CNNs on electrocardiogram signals. Aras et al [23] reported AUC = 0.89 for PH detection and AUC = 0.88 for PAH subtype classification. Finally, imaging-based methods apply transfer learning on radiographs. Kıvrak et al [24] trained MobileNetV2 on 6,642 chest X-rays and achieved 86.1% accuracy and AUC = 0.945. Jimenez-del-Toro et al [25] used CT texture analysis to detect CTEPH with 76% accuracy. Nazir et al [26] predicted PH risk in patients with COVID-19 from HRCT with 86.1% accuracy.

The current European Society of Cardiology (ESC)/European Respiratory Society (ERS) Guidelines (2022) define PH as mean pulmonary artery pressure (mPAP) >20 mmHg. Our study used the older threshold of mPAP ≥25 mmHg. This choice may bias the cohort toward more advanced disease and limit detection of early or borderline PH.

### 1.1. Motivation and novel contributions

Hemodynamic measurements with invasive RHC constitute the reference standard for diagnosing PH and categorizing PH into defined subtypes for appropriate downstream management: a task that cannot be replaced by less risky and expensive noninvasive imaging tests [27,28]. This creates a critical need for accurate, non-invasive alternatives that could potentially reduce the need for invasive procedures while maintaining diagnostic accuracy. For this reason, we were motivated to develop an efficient and accurate deep learning model for classifying PH—with reference graded hemodynamic severity as assessed using invasive RHC—using chest X-ray images. Therefore, we present ExMobileNet, a deep feature engineering model integrating multiple components. The model extracts features using MobileNetV2 [24], then selects key features through neighborhood component analysis (NCA) and chi-square (Chi2) feature selectors. Classification is performed using k-nearest neighbors (kNN) and support vector machine (SVM), with final decisions made through iterative voting and greedy algorithm fusion.

The contributions of this deep learning model (ExMobileNet) are: (1) A novel explainable AI (XAI) and deep feature engineering model for non-invasive PH severity classification; (2) A self-organized feature engineering pipeline combining transfer learning and traditional ML; (3) This study is among the first to apply deep feature engineering for automated PH detection from chest X-ray images.

This work addresses the clinical question of whether standard chest X-rays can replace invasive measures for PH severity stratification.

## 2. Methods

### 2.1. Dataset

We retrospectively collected 1,293 de-identified chest X-ray images (224 × 224 pixels) from Firat University Cardiology Clinic, Turkey, between January 1, 2016 and December 31, 2022. Chest X-ray images were acquired as close as possible to the RHC date. For 112 of 135 patients with PH, X-rays were taken on the same day as RHC. The remaining 23 patients had their images within 30 days before or after the procedure (median interval = 5 days). We selected the image nearest in time when multiple studies were available.

Patients with PH ($n$ = 135; age 34.12 ± 10.18 years, range 22–60):

- Inclusion: age 18–80 years and definitive PH diagnosis via invasive RHC (mean pulmonary artery pressure ≥25 mmHg) within 30 days of X-ray.
- Exclusion: inadequate image quality, history of thoracic surgery, and concomitant pulmonary disease unrelated to PH.

Healthy controls ($n$ = 551; age 52.07 ± 8.6 years, range 25–79):

- Inclusion: age 18–80 years, no history of cardiac or pulmonary disease, and availability of standard chest X-ray.
- Exclusion: any clinical or imaging evidence of cardiopulmonary pathology.

All patients with PH underwent RHC, measuring seven hemodynamic parameters: (1) cardiac output, (2) cardiac index, (3) pulmonary vascular resistance, (4) right atrial pressure, (5) mean pulmonary artery pressure, (6) systolic pulmonary artery pressure, and (7) diastolic pulmonary artery pressure. From these data, we created seven hemodynamic-specific datasets: (1) CO-Set, (2) CI-Set, (3) PVR-Set, (4) RAP-Set, (5) MPAP-Set, (6) SPAP-Set, and (7) DPAP-Set, each with severity-graded classes (Table 1).

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Fırat University institutional review (Approval No. E-2022-321; December 15, 2022). Written informed consent was obtained from all participants prior to inclusion.

To ensure robust model validation, we implemented a patient-level cross-validation strategy. As each patient with PH contributed multiple X-rays, we ensured that all images from the same patient were assigned to the same fold during cross-validation. This approach prevented data leakage and provided a more realistic assessment of model performance. The demographic and clinical characteristics of the study population are summarized in Table 2. All patients were referred for RHC owing to suspected PH. We did not include asymptomatic or mild cases. Consequently, the model reflects a preselected clinical population rather than a general screening cohort.

### 2.2. Proposed ExMobileNet model

The model architecture comprised four phases: (1) feature extraction: raw chest X-ray images (224 × 224) are divided into nine overlapping patches (112 × 112). The original image and patches are processed through MobileNetV2's [29] fully-connected (FC) and global average pooling (GAP) layers, generating feature vectors ($f$) that are concatenated into two feature vectors ($X_1$ and $X_2$), and further combined into a third vector ($X_3$); (2) feature selection: NCA [30] and Chi2 [31] functions process all three feature vectors, producing six selected feature vectors ($s$) with reduced dimensionality; (3) classification: kNN [32] and SVM [33,34] classifiers analyze the selected features, generating 12 classifier-wise results ($r$); and (4) information fusion: an IMV [35] operation processes the classifier results, producing 10 additional voted results ($v$). A greedy algorithm [36] selects the highest-accuracy result from all 22 outcomes ($12\,r + 10\,v$) as the final model output (Figure 1). The mathematical operations for the proposed ExMobileNet are detailed in seven steps given below.

#### 2.2.1. Feature extraction

The FC and GAP layers of the pre-trained MobileNetV2 [29], trained on the ImageNet1K [37] dataset, were used to extract features from the input raw chest X-ray images and derived fixed-size patches.

Step 1 : Creation of nine overlapping patches (112 × 112) from each input image (224 × 224) using a stride of 56 × 56 (Figure 2).

$$patch_{cnt} = Im(i : i + ii, j : j + jj, :)$$
$$where\ cnt \in \{1, 2, \ldots, 9\};\ i, j \in \{1, 57, \ldots, 113\};\ ii, jj \in \{1, 2, \ldots, 111\} \quad (1)$$

where $patch$ represents fixed-sized patch and $Im$ is the raw image.

Step 2 : Extraction of features from the raw image and generated fixed-size patches using the FC and GAP layers of MobileNetV2.

**Table 1** Study datasets with patient classes determined by grades of hemodynamic severity

| Dataset names | Classes | Images by class (*n*) | Total images[a] (*n*) |
|---|---|---|---|
| D1: CO | 1: Cardiac output ≥4 l/min | 1: 487 | 1,290 |
| | 2: Cardiac output <4 l/min | 2: 252 | |
| | 3: Healthy[b] | 3: 551 | |
| D2: CI | 1: Cardiac index <2.5 l/min/m$^2$ | 1: 322 | 1,283 |
| | 2: Cardiac index ≥2.5 l/min/m$^2$ | 2: 410 | |
| | 3: Healthy | 3: 551 | |
| D3: PVR | 1: 2 ≤ PVR <5 Woods unit | 1: 158 | 1,291 |
| | 2: 5 ≤ PVR <7 Woods unit | 2: 173 | |
| | 3: PVR ≥7 Woods unit | 3: 409 | |
| | 4: Healthy | 4: 551 | |
| D4: RAP | 1: 5 mmHg ≤ RAP <10 mmHg | 1: 202 | 1,154 |
| | 2: 10 mmHg ≤ RAP <15 mmHg | 2: 321 | |
| | 3: RAP ≥15 mmHg | 3: 80 | |
| | 4: Healthy | 4: 551 | |
| D5: MPAP | 1: 25 mmHg ≤ MPAP <35 mmHg[c] | 1: 177 | 1,293 |
| | 2: 35 mmHg ≤ MPAP <45 mmHg | 2: 162 | |
| | 3: MPAP ≥45 mmHg | 3: 403 | |
| | 4: Healthy | 4: 551 | |
| D6: SPAP | 1: SPAP <70 mmHg | 1: 280 | 1,293 |
| | 2: SPAP ≥70 mmHg | 2: 462 | |
| | 3: Healthy | 3: 551 | |
| D7: DPAP | 1: DPAP <20 mmHg | 1: 213 | 1,293 |
| | 2: DPAP ≥20 mmHg | 2: 529 | |
| | 3: Healthy | 3: 551 | |

D: Dataset; CO: cardiac output; CI: cardiac index; PVR: pulmonary vascular resistance; RAP: right atrial pressure; MPAP: mean pulmonary artery pressure; SPAP: systolic pulmonary artery pressure; DPAP: diastolic pulmonary artery pressure.

[a] All datasets were derived from the collected 1,293 images; the total number of images in some datasets were <1,293 as some measurements were unavailable in patients with pulmonary hypertension.

[b] Healthy controls did not undergo invasive right heart catheterization, and hemodynamic measurements were not obtained; their number of images (551) remained constant across the datasets.

[c] For pulmonary hypertension diagnosis, mean pulmonary artery pressure must be ≥25 mmHg.

**Table 2** Demographic and baseline characteristics of the study population

| Characteristic | Patients with PH (*n* = 135) | Healthy Controls (*n* = 551) |
|---|---|---|
| Age (mean ± SD (range), years) | 34.12 ± 10.18 (22–60) | 52.07 ± 8.6 (25–79) |
| Sex (M / F, *n*) | 78 / 57 | 298 / 253 |
| mPAP (mean ± SD, mmHg) | 42.3 ± 9.8 | — |
| PVR (mean ± SD, woods units) | 5.8 ± 2.1 | — |
| CI (mean ± SD, l/min/m$^2$) | 2.1 ± 0.6 | — |
| RAP (mean ± SD, mmHg) | 9.4 ± 4.3 | — |
| SPAP (mean ± SD, mmHg) | 78.5 ± 18.2 | — |
| DPAP (mean ± SD, mmHg) | 24.7 ± 6.1 | — |
| Underlying PH subtypes | PAH: 62 (46%), CTEPH: 41 (30%), and Others: 32 (24%) | — |

mPAP: mean pulmonary artery pressure; PVR: pulmonary vascular resistance; CI: cardiac index; RAP: right atrial pressure; SPAP: systolic pulmonary artery pressure; DPAP: diastolic pulmonary artery pressure; PH: pulmonary hypertension; PAH: pulmonary arterial hypertension; CTEPH: chronic thromboem bolic pulmonary hypertension.

$$f_1^t = Mv2\left(Im, layer_t\right),\ t \in \{1,2\}\ layer \in \{FC, GAP\}$$
$$f_{k+1}^t = Mv2\left(patch_k, layer_t\right)\ k \in \{1,2,\ldots,9\} \quad (2)$$

where $f$ represents feature vector, and $Mv2(.,.)$, the pre-trained MobileNetV2. FC extracted 10 feature vectors ($f^1$) each of length 1000; and GAP, 10 feature vectors ($f^2$) each of length 1280.

*Step 3 :* Generation of final feature vectors by concatenating the extracted feature vectors.

$$X_t = concat\left(f_1^t, f_2^t, \ldots, f_{10}^t\right)$$
$$X_3 = concat\left(X_1, X_2\right) \quad (3)$$

where $X$ represents feature vectors. Three final feature vectors were generated: $X_1$, $X_2$, and $X_3$ of lengths 10,000 (=1000 × 10), 12,800 (=1280 × 10), and 22,800 (=10,000+12,800), respectively.

### 2.2.2. Feature selection

We used two used feature selectors: NCA and Chi2. NCA is a distance-based method that determines the weight of every feature and selects the features by ranking the weights. Chi2 is a fast statistical feature selection method. The parameter settings for the NCA selector were: stochastic gradient descent (as solver) and half the number of chest X-ray images (as number of iterations). For the Chi2 selector, we used the default settings. The general block diagram of the presented feature selection approximation is illustrated in Figure 3.

To better explain the recommended feature selection approximation, the steps have been given below.

*Step 4 :* Selection of the most informative 250 features from each of the final feature vectors ($X_1$, $X_2$, *and* $X_3$).

$$idx_{2g-1} = method\left(X_g, y\right)\ g \in \{1,2,3\},\ method \in \{NCA,\ Chi2\} \quad (4)$$

where $idx$ represents generated qualified index, $y$ is the real output, $NCA(.,.)$ and $Chi2(.,.)$, represents the NCA and Chi2 function, respectively. Six qualified index arrays were generated, which were used to select the most informative features from the final feature vectors.

$$s_{2g-1}(d,h) = X_g\left(d, idx_{2g-1}(h)\right)\ h \in \{1,2,\ldots,250\}, d \in \{1,2,\ldots,NI\}$$
$$s_{2g}(d,h) = X_g\left(d, idx_{2g}(h)\right) \quad (5)$$

where $s$ represents the selected feature vector of length 250, and $NI$ represents the number of images. Six selected feature vectors were generated, each containing the most information on 250 features.

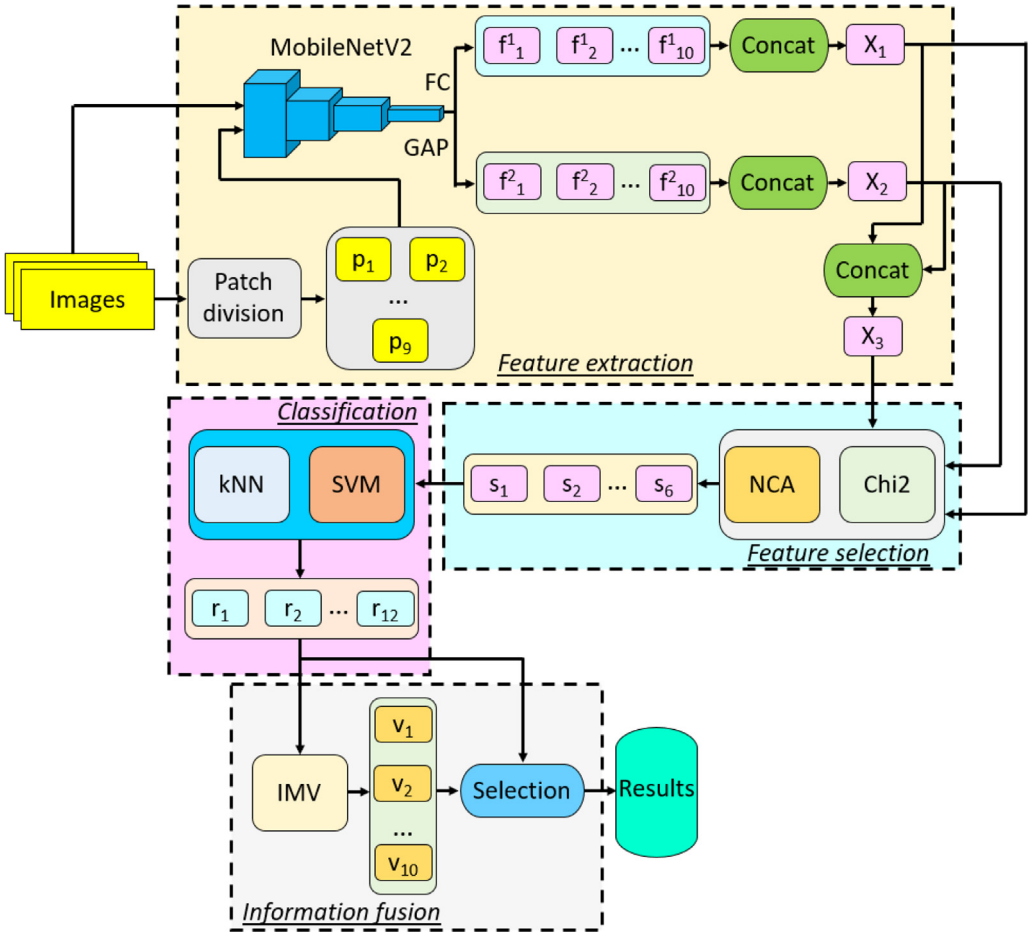



**Figure 1.** Schema of ExMobileNet model architecture. The workflow consists of: (1) Feature extraction, where chest X-ray images (224 × 224) are divided into patches (112 × 112) and processed through MobileNetV2's FC and GAP layers to generate feature vectors ($X_1$, $X_2$, $X_3$); (2) Feature selection using NCA and Chi2 functions to produce six reduced feature vectors; (3) Classification through kNN and SVM classifiers generating 12 results and (4) Information fusion using IMV and greedy algorithm to select the optimal final output from 22 total results (12 classifier + 10 voted results).
FC: fully-connected; GAP: global average pooling; NCA: neighborhood component analysis; Chi2: chi-square; kNN: k-nearest neighbors; SVM: support vector machine; IMV: integrated machine vision.

### 2.2.3. Classification

We employed two shallow classifiers based on preliminary performance testing: kNN and SVM. The kNN classifier was configured with k=10, squared inverse weights, L1-norm (Manhattan) distance, and 10-fold cross-validation. The SVM classifier used a third-order polynomial kernel, one-vs-one coding, C value of 1, and 10-fold cross-validation.

To validate performance, we applied stratified 10-fold cross-validation. The dataset was divided into 10 parts while maintaining equal class proportions across folds. For each iteration, nine parts were used for model training and parameter optimization, whereas one part was reserved for testing, ensuring every sample was used for testing once. For hyperparameter optimization, we performed grid search on the training data. For kNN, we tested k values from 1 to 20, with k = 10 showing optimal performance. SVM optimization explored complexity parameters (C = 0.1, 1, and 10) with a third-order polynomial kernel, with C = 1 selected as optimal. All parameter tuning occurred strictly within training folds to prevent data leakage. The graphical outline of this hybrid classification approximation has been showcased in Figure 4.

*Step 5 :* Calculation of classifier-wise results by inputting selected feature vectors to kNN and SVM classifiers.

$$r_q = classifier\left(X_q, y\right),\ q \in \{1, 2, \ldots, 6\},\ classifier \in \{kNN, SVM\} \tag{6}$$

This operation produces 12 results: $r_1$ to $r_6$ (kNN) and $r_7$ to $r_{12}$ (SVM).

### 2.2.4. Information fusion

The fusion process combines integrated machine vision (IMV) and the greedy algorithm approaches. The main objective of this research is to obtain the best classification performance and to use the utilized method's advantages together. The graphical outline (Figure 5) and steps of this phase is demonstrated in Figure 5.

Information fusion uses two stages: IMV and a greedy algorithm. IMV starts by taking the 12 individual classifier results as input. It then sorts these classifiers by their validation accuracy in descending order. The process continues by performing majority voting using the top c classifiers, where c ranges from 3 to 12. For each value of c, the system collects predictions from the top c classifiers and applies a mode function to find the most common class label, generating one voted outcome. This process produces 10 different voted outcomes (for c = 3 to 12). The greedy optimization stage combines all available results, including the 12 original classifier predictions and 10 voted outcomes from IMV, for a total of 22 results. It then evaluates the accuracy of all these results on validation data and selects the single result with the highest validation accuracy. This optimal selection is used for the final prediction.

*Step 6 :* Configuration of the IMV with classification accuracy as sorting criterion (descending order), iteration range from 3 to 12, and mode as majority function.

$$qi = \rho(cac)$$
$$cac_z = \Theta\left(r_z, y\right),\ z \in \{1, 2, \ldots, 12\} \tag{7}$$

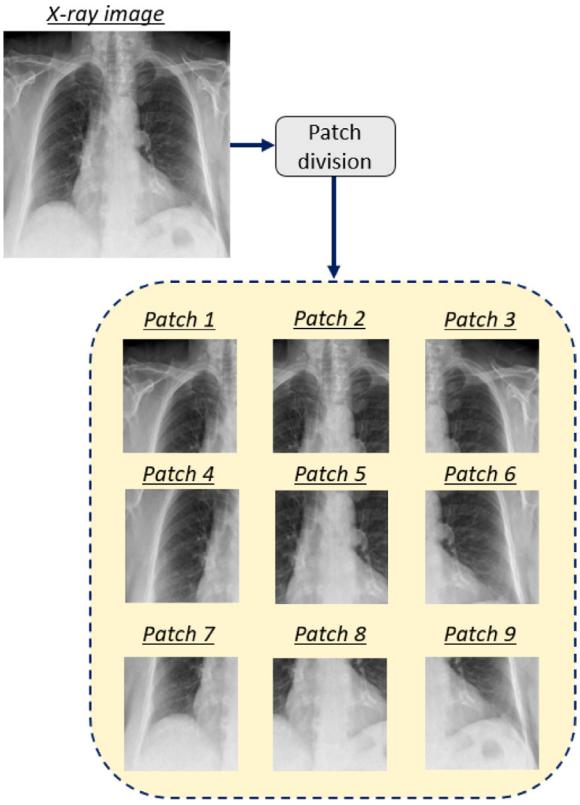



**Figure 2.** Patch division of raw 224 × 224 image into nine overlapping 112 × 112 patches.

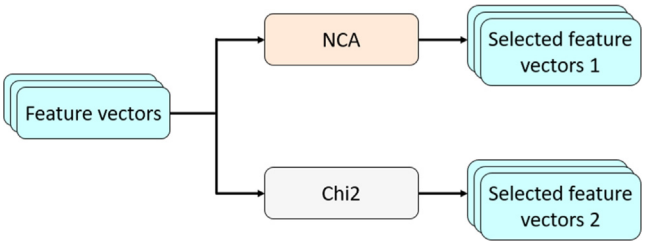



**Figure 3.** Presented feature selection approximation.
NCA: neighborhood component analysis; Chi2: chi-square.

where $cac$ represents classification accuracy; $\Theta(.)$ is classification accuracy computation function; $qi$ denotes qualified index of classifier-wise results, and $\rho(.)$ is the sorting function. The generated classifier-wise results were sorted in descending order of classification accuracies, then, the voted results ($v$) were generated using the qualified indexes, sorted classifier-wise results, and mode function.

$$v_{c-2} = \varpi\left(r_{qi(1)}, r_{qi(2)}, \ldots, r_{qi(c)}\right), \quad c \in \{3, 4, \ldots, 12\} \tag{8}$$

where $\varpi(.)$ represents mode function. The 10 (=12-3+1) voted results were generated based on the iteration range of 3 to 12.

*Step 7*: Selection of the result with the maximum classification accuracy from among the generated 22 results (12 classifier-wise results + 10 voted results).

$$final\_result = argmax\left(cac\left(\left[r_1, \ldots, r_{12}, v_1, \ldots, v_{10}\right]\right)\right) \tag{9}$$

These given seven steps were used to define the recommended image classification model.

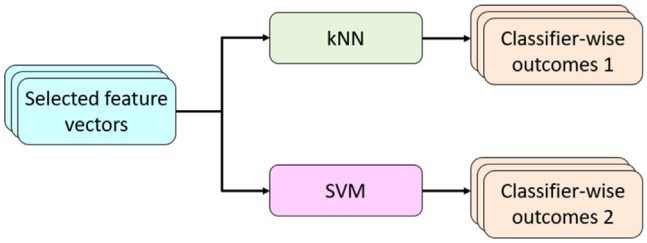



**Figure 4.** Presented classification approximation. Herein, two classifiers were used to generate classifier-wise outcomes.
kNN: k-nearest neighbors; SVM: support vector machine.

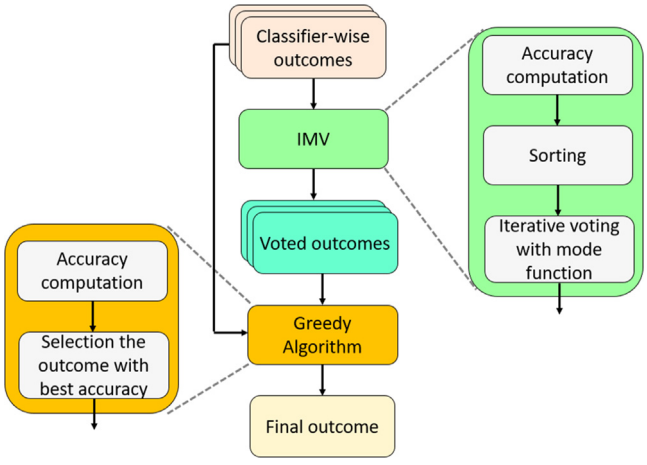



**Figure 5.** Outline of the recommended IMV and greedy-based information fusion approximation.
IMV: integrated machine vision.

## 3. Experimental results

We implemented ExMobileNet using MATLAB (2024a), incorporating pre-trained MobileNetV2, NCA, Chi2, kNN, SVM, and IMV components. Model performance was evaluated using two metrics: classification accuracy (standard benchmark) and geometric mean (more suitable for our mildly imbalanced datasets).

We used 10-fold cross-validation. For each metric, we reported the mean and standard deviation across all folds. We also computed 95% confidence intervals (CIs) by bootstrap resampling with 1000 iterations. To compare feature extraction methods (FC, GAP, and FC+GAP), we applied paired t-tests on the fold-level accuracy scores. We set statistical significance at $P < 0.05$.

### 3.1. Classifier-wise results

ExMobileNet generated 12 classifier-wise results for each of the seven datasets; the best accuracy was observed in Dataset 7 with the SVM classifier (Table 3). From the clinical standpoint, Dataset 5, which stratified chest X-ray images into non-PH and three PH classes with mean pulmonary artery pressures ≥25 mmHg (the reference diagnostic threshold)—is the most relevant, as it enables the prediction of RHC-assessed PH from non-invasive imaging input. Here, the best four-class classification accuracy and geometric mean were calculated by SVM to be 91.18% and 84.62%, respectively. Dataset 3—which stratified chest X-ray images into non-PH and three PH classes with different severities of elevated pulmonary vascular resistance ≥2 Woods unit—is interesting as the degree of pulmonary vascular resistance, a measure of pulmonary artery stiffness, correlates with PH severity, and also has implications for PH reversibility after correction of congenital heart defects (very high

**Table 3** Classifier-wise results for using SVM classifier (%)

| Dataset | Classification accuracy | | | Geometric mean | | |
|---|---|---|---|---|---|---|
| | Min. | Max. | Avg. | Min. | Max. | Avg. |
| D1: CO | 87.91 | 91.16 | 89.39 | 81.70 | 85.35 | 83.58 |
| D2: CI | 86.36 | 89.87 | 87.83 | 84.04 | 86.93 | 85.21 |
| D3: PVR | 84.43 | 87.45 | 86.02 | 75.93 | 78.26 | 77.08 |
| D4: RAP | 85.70 | 88.21 | 87.22 | 75.40 | 79.38 | 77.09 |
| D5: MPAP | 85.92 | 89.56 | 87.42 | 78.79 | 82.18 | 80.02 |
| D6: SPAP | 86.47 | 91.03 | 88.58 | 82.77 | 87.57 | 84.87 |
| D7: DPAP | 89.17 | 94.20 | 91.45 | 82.89 | 89.12 | 86.10 |

D: Dataset; CO: cardiac output; CI: cardiac index; PVR: pulmonary vascular resistance; RAP: right atrial pressure; MPAP: mean pulmonary artery pressure; SPAP: systolic pulmonary artery pressure; DPAP: diastolic pulmonary artery pressure.

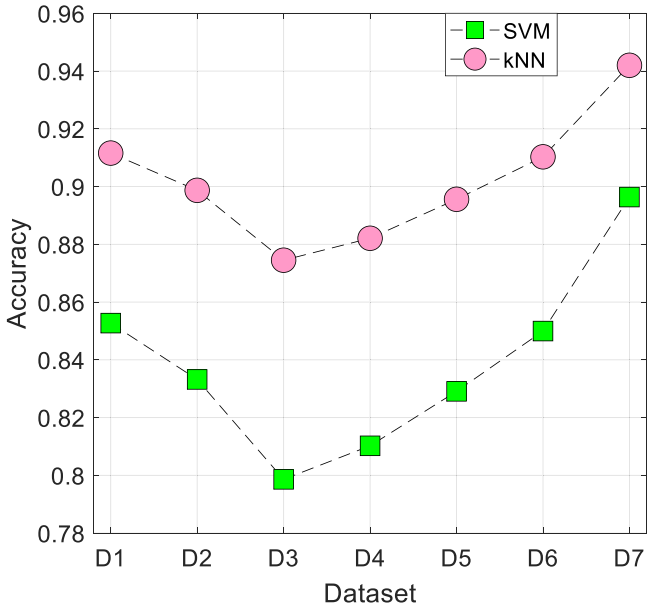



**Figure 6.** Benchmark of the used kNN and SVM classifier.
kNN: k-nearest neighbors; SVM: support vector machine.

PVR is a contraindication for surgery). Here, the best four-class classification accuracy and geometric mean were calculated by SVM to be 89.16% and 80.58%, respectively. Dataset 4—which stratified chest X-ray images into non-PH, one PH class with normal right atrial pressure <10 mmHg, and two PH classes with right atrial pressure ≥10 mmHg (the threshold for decompensated PH)—is important decompensated PH portends poor prognosis and signifies the need for escalation of treatment. Here, the best four-class classification accuracy and geometric mean were calculated by SVM to be 89.60% and 79.90%, respectively.

Moreover, kNN and SVM classifiers' classification accuracies on these datasets have been illustrated in Figure 6.

### 3.2. Voted results

ExMobileNet generated 10 IMV-voted results for each of the seven datasets; the results were generally numerically better than the classifier-wise results, and the best accuracy was observed in Dataset 7 (Table 3). For Dataset 5, the best four-class classification accuracy and geometric mean were slightly improved to 91.80% and 86.18%, respectively. For Dataset 3, the best four-class classification accuracy and geometric mean were also slightly improved to 90.55% and 83.34%, respectively. The same was true for Dataset 4, where the best four-class classification accuracy and geometric mean were 90.81% and 82.06%, respectively. The performance ranges of IMV-voted results across the seven datasets are given in Table 4.

### 3.3. Final model results

ExMobileNet achieved classification accuracies >90% and geometric mean values >80% across all seven datasets (Table 4). The confusion matrixes are shown in Figure 4.

For Dataset 5, when merging the three PH classes into a single class, the model demonstrated excellent binary classification performance for RHC-assessed PH versus non-PH, with sensitivity of 99.33% and specificity of 99.27%.

Similarly, for Dataset 3, the binary classification of RHC-assessed elevated PVR versus non-PH showed excellent performance with 98.51% sensitivity and 99.46% specificity. When combining the two PH classes with RAP >10 mmHg and grouping the remaining classes, the model achieved good performance in distinguishing decompensated PH from non-decompensated PH or non-PH (sensitivity: 96.75% and specificity: 95.56%). This capability serves as a valuable complement to clinical assessment in symptomatic patients with PH. The computed confusion matrices have been demonstrated in Figure 7 and classification results computed deploying Figure 7 have been listed in Table 5.

Moreover, receiver operating characteristic (ROC) curves for all datasets were created and are depicted in Figure 8. The values for the seven datasets are 94.07%, 93.80%, 71.45%, 89.75%, 92.50%, 94.10%, and 94.33 %, respectively.

Statistical analysis demonstrated significant differences between feature extraction methods. One-way ANOVA comparing FC, GAP, and FC+GAP feature vectors yielded $F = 58.23$, $P < 0.0001$. Post-hoc Bonferroni-corrected comparisons showed significant differences between all pairs: FC *vs*. GAP ($P = 0.0009$), FC *vs*. FC+GAP ($P = 0.0003$), and GAP *vs*. FC+GAP ($P = 0.0002$). Further ANOVA testing on NCA-selected vectors ($F = 46.71$, $P < 0.0001$) and Chi2-selected vectors ($F = 39.54$, $P < 0.0001$) confirmed distinct information capture, with all pairwise comparisons yielding $P < 0.01$. These results validate that each feature selection method contributes unique discriminative power to the model.

### 3.4. Ablation study

To evaluate the effectiveness of different feature extraction approaches, we conducted an ablation study comparing three methods: FC layers, GAP, and their combination (FC+GAP). The results demonstrated the relative performance of each approach across all seven datasets.

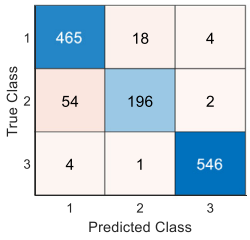
The FC feature extractor consistently outperformed other methods (Figure 9), achieving the highest average classification accuracy across most datasets (Table 6). The combination of FC+GAP did not yield additional benefits, suggesting that the increased model complexity did not translate to improved feature extraction. This finding indicates that FC layers alone are sufficient for optimal feature extraction in this context.

The ablation study also examined the impact of different feature selection methods and classifiers on model performance. In the feature selection phase, we compared two feature selection methods: NCA and Chi2. NCA demonstrated superior performance across all datasets (Figure 10), consistently selecting more informative features for classi-

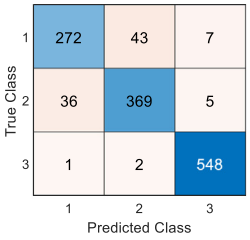
**Table 4** Performance ranges of IMV-voted results across the seven datasets (%)

| Dataset | Classification accuracy (%) | | | Geometric mean (%) | | |
|---|---|---|---|---|---|---|
| | Min. | Max. | Avg. | Min. | Max. | Avg. |
| D1: CO | 92.02 | 93.57 | 92.59 | 86.59 | 90.28 | 87.96 |
| D2: CI | 91.43 | 92.67 | 91.73 | 89.75 | 91.10 | 90.17 |
| D3: PVR | 89.62 | 90.55 | 90.11 | 80.70 | 83.34 | 81.90 |
| D4: RAP | 89.77 | 90.81 | 90.35 | 80.27 | 82.06 | 81.37 |
| D5: MPAP | 90.41 | 91.80 | 91.11 | 83.27 | 86.18 | 84.57 |
| D6: SPAP | 91.42 | 93.12 | 92.45 | 88.23 | 91.26 | 89.83 |
| D7: DPAP | 94.04 | 94.82 | 94.45 | 89.28 | 91.00 | 90.01 |

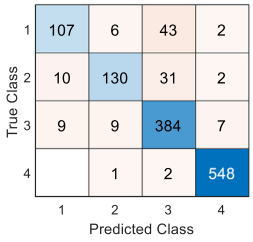
IMV: integrated machine vision; D: Dataset; CO: cardiac output; CI: cardiac index; PVR: pulmonary vascular resistance; RAP: right atrial pressure; MPAP: mean pulmonary artery pressure; SPAP: systolic pulmonary artery pressure; DPAP: diastolic pulmonary artery pressure.



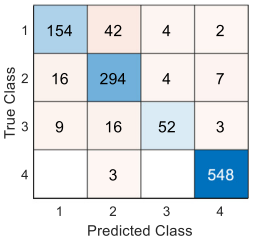

D1. 1: Cardiac output ≥4 l/min; 2: Cardiac output <4 l/min; 3: Healthy.



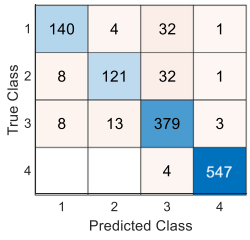

D2. 1: Cardiac index <2.5 l/min/m$^2$; 2: Cardiac index ≥2.5 l/min/m$^2$; 3: Healthy.



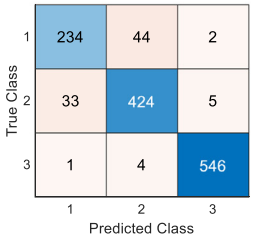

D3. 1: 2 WU ≤ PVR <5 WU; 2: 5 WU ≤ PVR <7 WU; 3: PVR ≥7 WU; 4: Healthy.



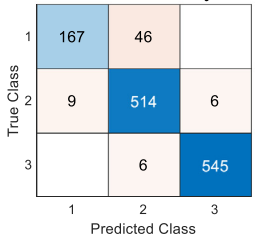

D4. 1: 5 mmHg ≤ RAP <10 mmHg; 2: 10 mmHg ≤ MPAP <15 mmHg; 3: MPAP ≥15 mmHg; 4: Healthy.




D5. 1: 25 mmHg ≤ MPAP <35 mmHg; 2: 35 mmHg ≤ MPAP <45 mmHg; 3: MPAP ≥45 mmHg; 4: Healthy.




D6. 1: SPAP <70 mmHg; 2: SPAP ≥70 mmHg; 3: Healthy.




D7. 1: DPAP <20 mmHg; 2: DPAP ≥20 mmHg; 3: Healthy.

**Figure 7.** Confusion matrixes of ExMobileNet for the seven datasets: D1. CO; D2. CI; D3. PVR; D4. RAP; D5. MPAP; D6. SPAP; D7: DPAP.
CO: cardiac output; CI: cardiac index; PVR: pulmonary vascular resistance; RAP: right atrial pressure; MPAP: mean pulmonary artery pressure; SPAP: systolic pulmonary artery pressure; DPAP: diastolic pulmonary artery pressure.

**Table 5** Final model results for the seven datasets (%)

| Dataset | Classification accuracy (%) | Geometric mean (%) |
|---|---|---|
| D1: CO | 93.57 | 90.28 |
| D2: CI | 92.67 | 91.10 |
| D3: PVR | 90.55 | 83.03 |
| D4: RAP | 90.81 | 81.97 |
| D5: MPAP | 91.80 | 86.18 |
| D6: SPAP | 93.12 | 91.26 |
| D7: DPAP | 94.82 | 91.00 |

D: Dataset; CO: cardiac output; CI: cardiac index; PVR: pulmonary vascular resistance; RAP: right atrial pressure; MPAP: mean pulmonary artery pressure; SPAP: systolic pulmonary artery pressure; DPAP: diastolic pulmonary artery pressure.

**Table 6** Average classification accuracies by feature generation method for the seven datasets (%)

| Dataset | FC (%) | GAP (%) | FC+GAP (%) |
|---|---|---|---|
| D1: CO | 90.52 | 89.42 | 90.35 |
| D2: CI | 88.87 | 87.59 | 88.62 |
| D3: PVR | 87.01 | 87.09 | 86.37 |
| D4: RAP | 88.19 | 87.67 | 87.76 |
| D5: MPAP | 88.54 | 87.90 | 88.19 |
| D6: SPAP | 89.75 | 89.35 | 89.06 |
| D7: DPAP | 92.78 | 92.11 | 92.11 |

FC: fully-connected; GAP: global average pooling; D: Dataset; CO: cardiac output; CI: cardiac index; PVR: pulmonary vascular resistance; RAP: right atrial pressure; MPAP: mean pulmonary artery pressure; SPAP: systolic pulmonary artery pressure; DPAP: diastolic pulmonary artery pressure.

fication. This superiority can be attributed to the ability of NCA to learn feature weights that maximize the expected leave-one-out classification accuracy. Furthermore, NCA considers feature interactions rather than evaluating features independently, and adapts to the underlying data distribution more effectively.

In the classification phase, we evaluated two standard shallow classifiers: kNN and SVM. SVM consistently outperformed kNN across all datasets (Figure 11), likely owing to its superior handling of high-dimensional feature spaces and better generalization capability. Additionally, SVM demonstrated robust performance in the presence of noise and effective management of non-linear relationships through kernel functions. The performance advantage of SVM was the most pronounced in complex multi-class scenarios (Datasets 3, 4, and 5) and cases with potentially overlapping class boundaries.

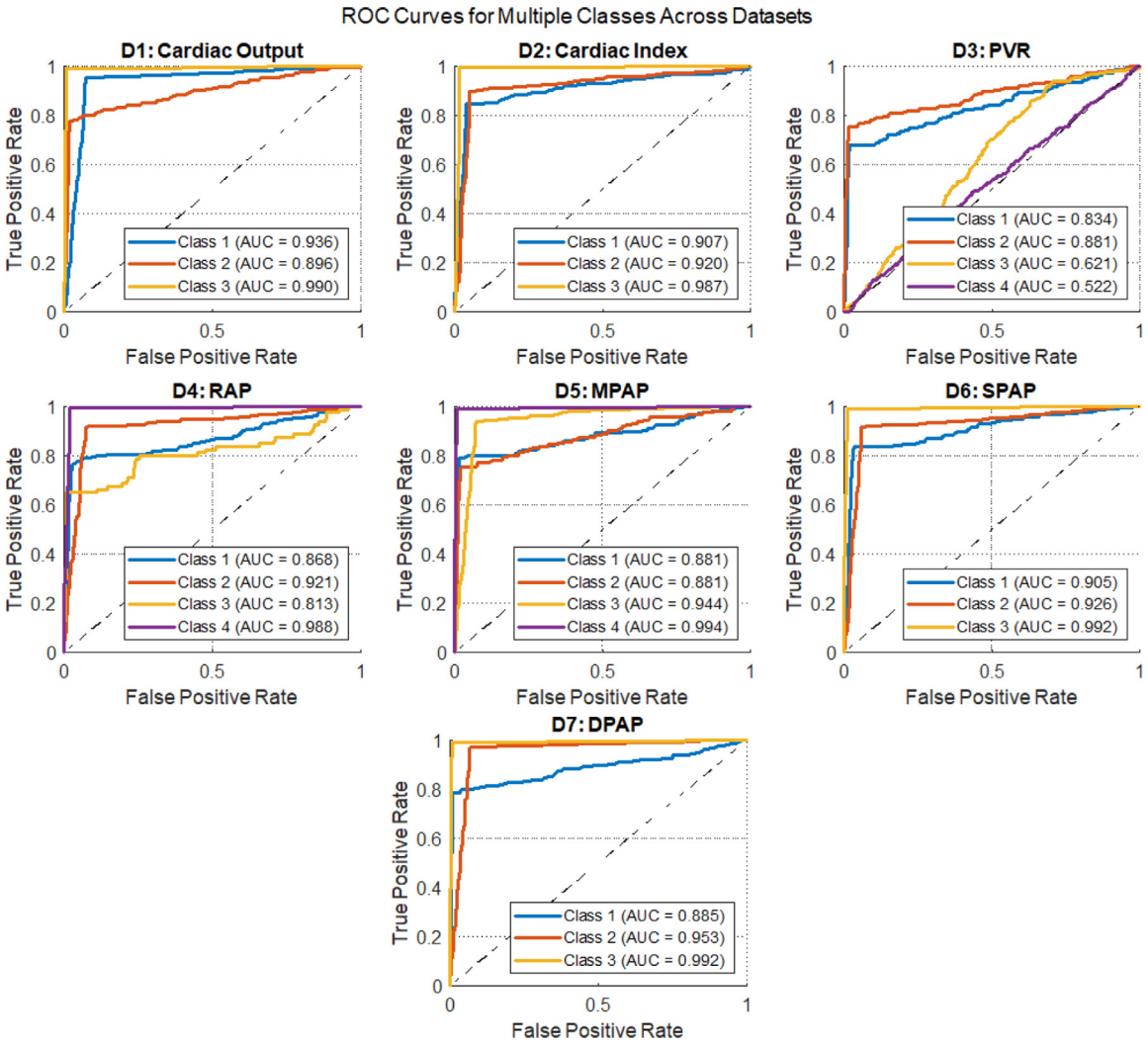



**Figure 8.** Computed ROC curves of the used seven datasets: D1. CO; D2. CI; D3. PVR; D4. RAP; D5. MPAP; D6. SPAP; D7: DPAP.
CO: cardiac output; CI: cardiac index; PVR: pulmonary vascular resistance; RAP: right atrial pressure; MPAP: mean pulmonary artery pressure; SPAP: systolic pulmonary artery pressure; DPAP: diastolic pulmonary artery pressure; ROC: receiver operating characteristic; AUC: area under the curve.

According to the results of this ablation results, the findings are: (1) The FC + NCA + SVM path matches full ExMobileNet in average accuracy; (2) Voting and fusion add robustness. They reduce performance drops in difficult cases; (3) Fusion selects the best result among 22 options. This guards against single-classifier failures; (4) Voting outputs provide insight into classifier agreement. This aids interpretability; (5) The extra steps add minimal runtime (<0.5 s per image). This is acceptable for clinical use; (6) Fusion and voting require no additional data. They reuse existing classifier outputs; (7) Overall, the complexity is justified by greater stability and clearer decision support.

### *3.5. Comparison with state-of-the-art*

In this section, we have compared to our classification results to (1) Swin Transformer [38], (2) Vision Transformer (ViT) [39], (3) ConvNeXt Tiny [40], and (4) ResNet50 [41]. The results are shown in Table 7.

Although the accuracy gains are small, ExMobileNet offers clear practical benefits. Its MobileNetV2 backbone has only 3.4 million parameters (14 MB), whereas ResNet50 has 25.6 million (98 MB), ViT-Base has 86 million (340 MB), and Swin Tiny has 28 million (112 MB).

## 4. Discussions

PH is a severe condition requiring invasive right heart catheterization for definitive diagnosis. Early, non-invasive screening tools could substantially impact patient care by enabling earlier detection and monitoring. Chest X-rays, being widely available and cost-effective, represent an ideal input for automated diagnostic support systems.

Our main motivation was to develop a new generation of highly accurate feature engineering models for biomedical image classification. ExMobileNet combines computer vision, transfer deep learning, and handcrafted feature engineering in a novel hybrid approach. The model is parametric, with design components in four main phases: (1) feature extraction, (2) feature selection, (3) classification, and (4) information fusion.

This study showed that ExMobileNet can classify PH severity from chest X-rays with high accuracy. It achieved >90% accuracy and >80% geometric mean score across seven hemodynamic tasks. The model performed best on mean pulmonary artery pressure and pulmonary vascular resistance. Information fusion improved results beyond any single classifier. Feature importance tools confirmed that the model focused on key anatomical regions.

Most chest X-rays were obtained on the same day as RHC, minimizing temporal bias. However, 17% of images fell outside this window

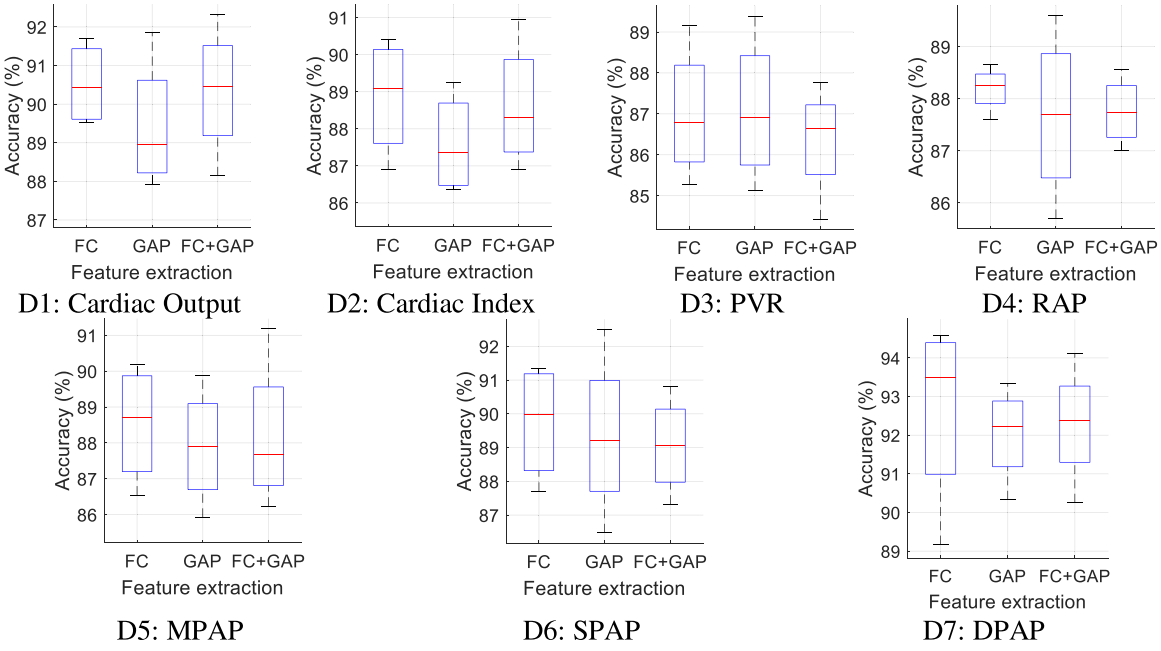



**Figure 9.** Classification accuracies by feature extraction method for the seven datasets: D1. CO; D2. CI; D3. PVR; D4. RAP; D5. MPAP; D6. SPAP; D7: DPAP. CO: cardiac output; CI: cardiac index; PVR: pulmonary vascular resistance; RAP: right atrial pressure; MPAP: mean pulmonary artery pressure; SPAP: systolic pulmonary artery pressure; DPAP: diastolic pulmonary artery pressure; FC: fully-connected; GAP: global average pooling.

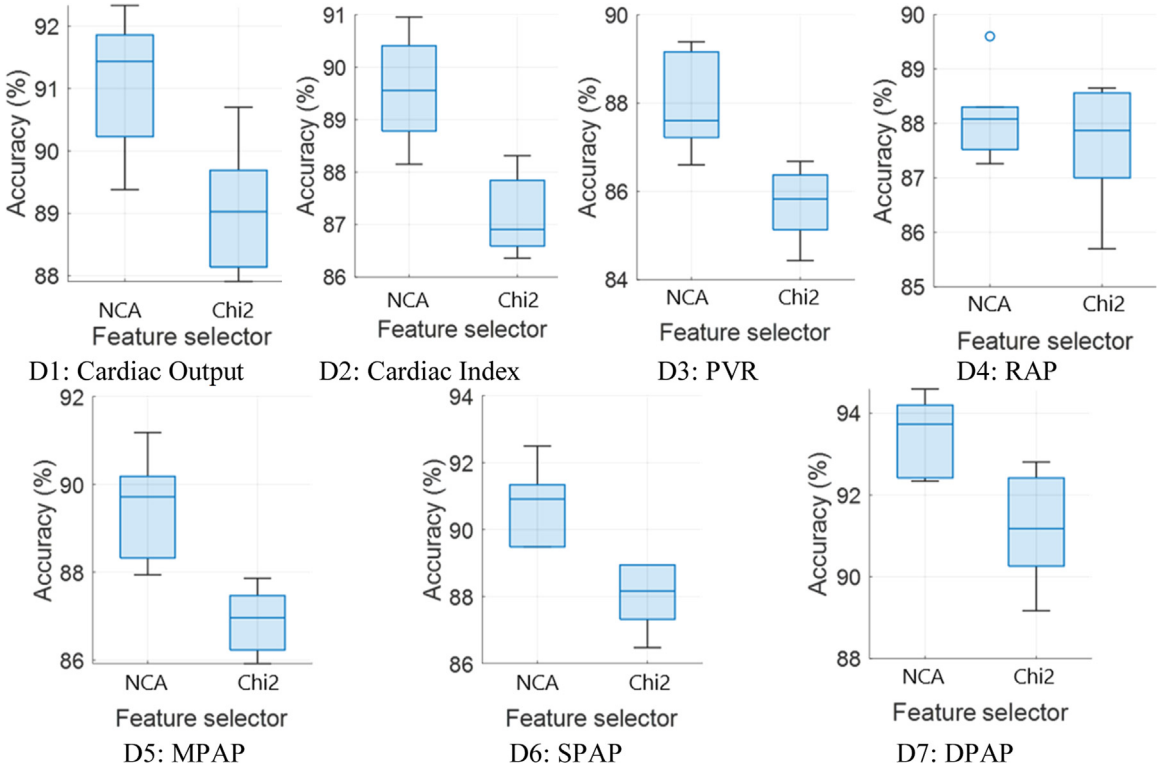



**Figure 10.** Classification accuracies by feature extraction method for the seven datasets: D1. CO; D2. CI; D3. PVR; D4. RAP; D5. MPAP; D6. SPAP; D7: DPAP. CO: cardiac output; CI: cardiac index; PVR: pulmonary vascular resistance; RAP: right atrial pressure; MPAP: mean pulmonary artery pressure; SPAP: systolic pulmonary artery pressure; DPAP: diastolic pulmonary artery pressure; NCA: neighborhood component analysis; Chi2: chi-square.

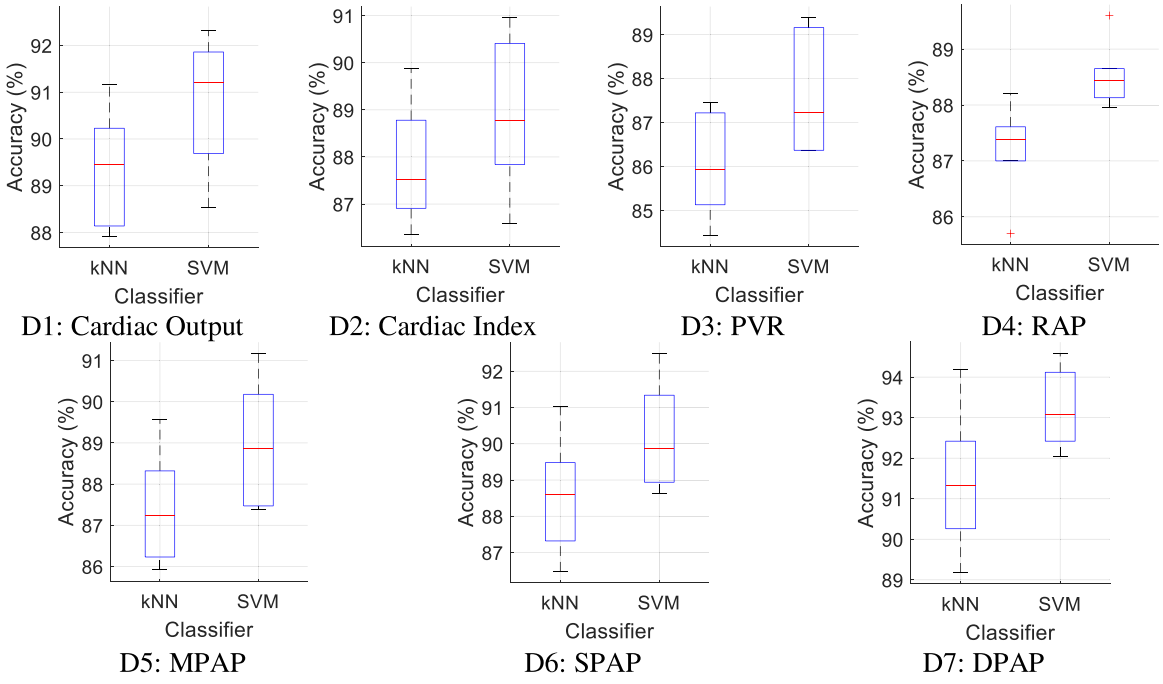



**Figure 11.** Classification accuracies by classifier for the seven datasets: D1. CO; D2. CI; D3. PVR; D4. RAP; D5. MPAP; D6. SPAP; D7: DPAP.
CO: cardiac output; CI: cardiac index; PVR: pulmonary vascular resistance; RAP: right atrial pressure; MPAP: mean pulmonary artery pressure; SPAP: systolic pulmonary artery pressure; DPAP: diastolic pulmonary artery pressure; kNN: k-nearest neighbors; SVM: support vector machine.

**Table 7** Benchmark classification accuracies (%)

| Dataset | Swin Transformer | ViT | ConvNeXt | ResNet50 | ExMobileNet (our model) |
|---|---|---|---|---|---|
| D1: CO | 92.79 | 92.64 | 92.48 | 92.25 | 93.57 |
| D2: CI | 91.89 | 91.74 | 91.58 | 91.35 | 92.67 |
| D3: PVR | 89.78 | 89.62 | 89.47 | 89.23 | 90.55 |
| D4: RAP | 90.03 | 89.95 | 89.69 | 89.51 | 90.81 |
| D5: MPAP | 91.03 | 90.87 | 90.72 | 90.49 | 91.80 |
| D6: SPAP | 92.34 | 92.19 | 92.03 | 91.80 | 93.12 |
| D7: DPAP | 94.04 | 93.89 | 93.74 | 93.50 | 94.82 |

D: Dataset; CO: cardiac output; CI: cardiac index; PVR: pulmonary vascular resistance; RAP: right atrial pressure; MPAP: mean pulmonary artery pressure; SPAP: systolic pulmonary artery pressure; DPAP: diastolic pulmonary artery pressure.

by up to 30 days. Changes in hemodynamics over that interval could alter the appearance of key features and affect model performance. To strengthen validity, future work should use same-day imaging and RHC or prospectively schedule both procedures within a narrow time frame. This alignment will ensure that the model predictions reflect true contemporaneous physiology.

ExMobileNet consistently achieved high classification accuracy (>90%) and strong geometric mean scores (>80%) across all seven datasets. For mean pulmonary artery pressure classification (Dataset 5), which directly relates to the diagnostic threshold for PH, the model achieved 91.80% accuracy in four-class classification. Notably, when considering binary classification (PH *vs*. non-PH), it demonstrated excellent sensitivity (99.33%) and specificity (99.27%), approaching the accuracy of invasive measurements.

In assessing pulmonary vascular resistance (Dataset 3), critical for determining treatment options and surgical candidacy, the model achieved 90.55% accuracy in distinguishing between different PVR severity levels. This capability could assist in early identification of patients requiring more aggressive intervention. For right atrial pressure assessment (Dataset 4), which indicates right heart decompensation, the model's 90.81% accuracy in identifying elevated RAP could help identify patients requiring urgent intervention.

### 4.1. Component analysis and model architecture

In the feature extraction phase, we used two feature extraction functions derived from the FC and GAP layers of MobileNetV2 to generate feature vectors; furthermore, a combined feature vector was constructed from the outputs of the two feature extractors (Figure 12).

In the information phase, the most accurate result was selected for each of the seven datasets from among the 12 classifier-wise and 10 IMV-voted results. By analyzing the parametric model components that contributed to the individual final results in all seven datasets, we observed that the FC layer, NCA, and SVM were the most commonly used for the feature extraction, feature selection, and classification phases, respectively, across all seven datasets. Across the seven datasets, the classification accuracies of the FC + NCA + SVM combination approximated the final results of ExMobileNet (Figure 13a) and were consistently ranked among the top 4 of the 12 classifier-wise results (Figure 13b).

### 4.2. Highlights

ExMobileNet introduces a novel deep feature engineering approach comprising four integrated phases: feature extraction, feature selection, classification, and information fusion. For feature extraction, the FC layer-based features yielded the best average classification accuracies

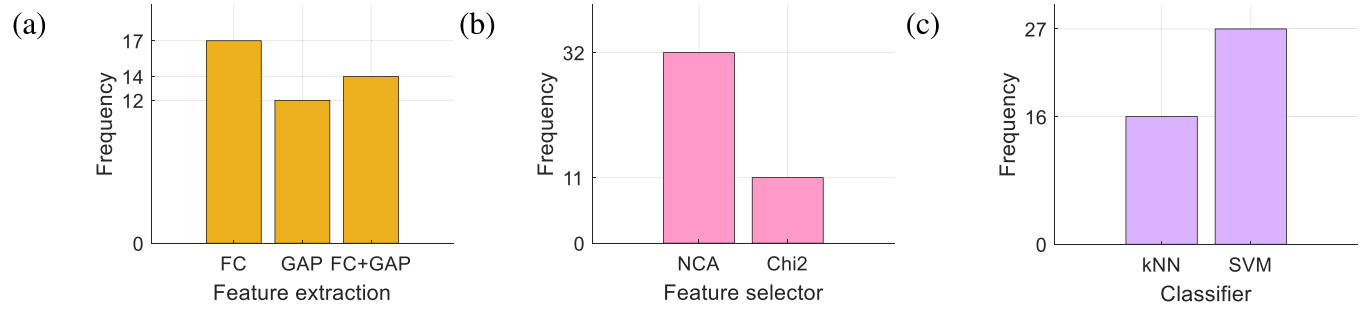



**Figure 12.** Histogram of the model components that contributed to the results across all seven datasets by (a) feature extractor, (b) feature selector, and (c) classifier.
FC: fully-connected; GAP: global average pooling; NCA: neighborhood component analysis; Chi2: chi-square; kNN: k-nearest neighbors; SVM: support vector machine.

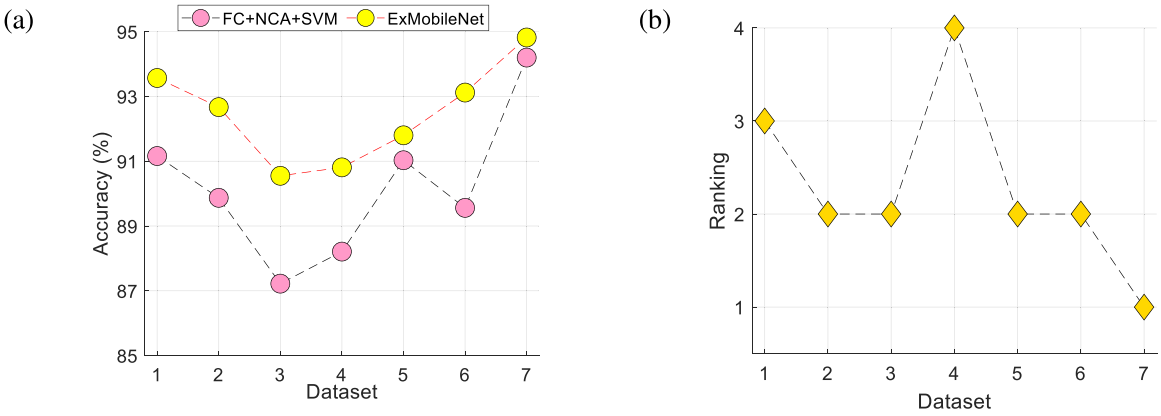



**Figure 13.** Comparison of the classification accuracy of (a) FC + NCA + SVM vs. ExMobileNet and (b) ranked against other classifier-wise results across the seven datasets.
FC: fully-connected; NCA: neighborhood component analysis; SVM: support vector machine.

for six out of seven datasets, outperforming GAP-based features and combination (FC+GAP)-based features. For feature selection, the NCA selector outperformed Chi2 in selecting the most informative features across the datasets. For classification, SVM consistently attained higher classification performance than kNN. For information fusion, the final model results were all derived from IMV-voted outcomes, which provided indirect support for the validity of our design choice. Overall, the FC + NCA + SVM combination emerged as particularly effective, consistently ranking among the top performer configurations across all datasets.

The ExMobileNet model offers the advantage of a well-rounded approach that integrates feature extraction, feature selection, classification, and information fusion into a single deep feature engineering model. Moreover, the self-organized architecture enables the automatic selection of the optimal results based on classification accuracies computed upstream. The design of ExMobileNet allows handling of different datasets effectively, as seen by its consistently high performance across the seven parallel chest X-ray image datasets. By analyzing the model components contributing to model accuracy using histograms, we can dissect and identify methods and combinations that worked best for individual datasets and globally, which confers transparency and insight into the model's operations. Importantly, the information fusion phase maximizes classification performance, making the model more robust and accurate in its predictions.

### 4.3. Potential implications

The potential implications of our ExMobileNet-based model include enabling early PH detection, reducing invasive catheterization needs, providing automated severity reports to radiologists, accelerating treatment decisions, lowering hospital costs, integration with existing X-ray systems, and potential adaptation to other cardiopulmonary diseases.

Although ExMobileNet demonstrated strong overall performance, understanding which image regions drive its predictions remains essential for clinical adoption. We conducted feature importance analysis using gradient-weighted class activation mapping (Grad-CAM) and SHapley Additive exPlanations (SHAP): (1) Grad-CAM heatmaps highlighted the pulmonary artery silhouette, right ventricular contour, and cardiothoracic ratio as key areas of focus; (2) SHAP values confirmed that these regions contributed the most to the model's output scores across all hemodynamic tasks.

This deeper interpretability supports clinicians in two ways. First, it shows that ExMobileNet attends to anatomically relevant structures rather than spurious patterns. Second, it builds trust by linking model decisions to features familiar to radiologists.

Future work will include comprehensive region-level importance evaluation on larger datasets, development of rule-based clinical guidelines, and user studies with radiologists to assess the impact on diagnostic confidence.

### 4.4. Limitations and future works

Although our study dataset provided comprehensive hemodynamic information, we acknowledge important limitations. The data collection from a single institution may restrict model generalizability and high-performance metrics could indicate data-specific overfitting or sampling bias. No independent external cohort was used for validation.

To address these limitations, our future work will focus on multi-institutional validation to ensure robust performance across different hospitals and imaging protocols, diverse patient populations, and

varied clinical settings. Additionally, we plan to collect larger multicentric datasets and evaluate alternative backbone networks beyond MobileNetV2. These steps will help establish the model's real-world clinical utility and generalizability.

In the future, we aim to evaluate alternative backbone networks, including recent transformer-based models and investigate more sophisticated fusion strategies. Additionally, we will explore optimizations to enable deployment in resource-constrained clinical settings without compromising performance. By using larger datasets and more precise deep learning models, our goal is to develop next-generation heart disease detection applications that may be implemented as clinician-assistance tools in real-world settings. Our model was trained and validated using the historical mPAP ≥25 mmHg criterion. Consequently, ExMobileNet currently detects established PH rather than early-stage or borderline cases. To address this limitation, future work should include patients with mPAP values between 21 and 24 mmHg. We will retrain and test the model using the ESC/ERS 2022 standard (mPAP >20 mmHg) to evaluate its ability to identify early PH. This step will help extend the clinical utility of ExMobileNet for earlier screening and intervention.

Our cohort comprised patients with clinical suspicion of PH who underwent RHC. We did not evaluate the model in asymptomatic individuals or those with mild or borderline disease. Therefore, ExMobileNet's performance as a general screening tool remains untested. Future work should enroll broader populations, including asymptomatic subjects and early-stage cases, to determine its utility for true PH screening.

Figure 7 reveals notable confusion between adjacent severity classes, particularly in the MPAP (D5) and RAP (D4) datasets. These misclassifications have considerable clinical implications, for instance, incorrectly classifying a moderate MPAP case as mild could delay necessary treatment escalation, whereas RAP misclassification may lead to suboptimal risk stratification. To address these limitations, our future work will focus on strengthening class separation through the incorporation of additional clinical markers and development of hierarchical classifiers to reduce adjacent-class confusion.

To adapt ExMobileNet for early-stage PH detection, we will retrain the model on a dataset that includes patients with mPAP values of 21–24 mmHg under the ESC/ERS 2022 standard. We will label these cases as mild PH and include them alongside healthy controls. The model will then learn features that distinguish subtle vascular changes. We will evaluate its performance on a binary task (mPAP >20 mmHg vs. ≤20 mmHg) and report sensitivity and specificity for mild cases. Finally, we will validate this version prospectively in an at-risk population to confirm its ability to detect early PH in real-world settings.

### 4.5. Summary

ExMobileNet demonstrated robust and high performance in classifying pulmonary hypertension severity from routine chest X-rays, achieving over 90% accuracy across seven hemodynamic parameters. Decision fusion improved results beyond individual classifiers, with the FC + NCA + SVM pathway consistently matching the final model performance. These results indicate that ExMobileNet can effectively support early PH detection and serve as a non-invasive complement to current diagnostic tools.

Future work will focus on three key areas: clinical validation, technical enhancement, and practical implementation. We plan to conduct prospective multi-center clinical studies to validate ExMobileNet across diverse patient populations and imaging devices. Technical improvements will enhance model explainability through user-friendly heatmaps and rule-based reports. For practical implementation, we will integrate the model into radiology workflows and assess its impact on diagnostic decision-making. These steps aim to establish the generalizability and clinical utility of ExMobileNet in real-world practice.

## Conflicts of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Fundings

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

## Author contributions

**Tarık Kıvrak:** Validation, Resources, Formal analysis, Methodology, Investigation, Writing – review & editing, Writing – original draft. **Mehmet Ali Gelen:** Writing – review & editing, Writing – original draft, Validation, Resources, Methodology, Investigation, Conceptualization. **Ozge Salkin:** Investigation, Formal analysis, Writing – review & editing, Writing – original draft, Validation, Resources. **Ozkan Karaca:** Writing – review & editing, Writing – original draft, Validation, Resources, Investigation, Formal analysis. **Prabal Datta Barua:** Writing – review & editing, Writing – original draft, Validation, Investigation, Formal analysis. **Sengul Dogan:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Conceptualization. **Turker Tuncer:** Writing – review & editing, Writing – original draft, Validation, Software, Resources, Methodology, Investigation, Conceptualization. **Ru-San Tan:** Writing – review & editing, Writing – original draft, Validation. **Massimo Salvi:** Writing – review & editing, Writing – original draft, Validation, Investigation. **Vinitha Sree Subbhuraam:** Writing – original draft, Validation, Writing – review & editing. **U.R. Acharya:** Writing – review & editing, Writing – original draft, Validation, Project administration.

## References

[1] Neppala S, Chigurupati HD, Fath A, et al. Impact of pulmonary hypertension on short and long-term outcome after mitral transcatheter edge-to-edge repair: A meta-analysis. Cardiovasc Revasc Med 2025. doi:10.1016/j.carrev.2024.12.012.
[2] Wander G, Montanaro C, Dixit P, et al. Pregnancy and pulmonary artery hypertension: Management challenges. Int J Cardiol Congenit Heart Dis 2025;19:100566. doi:10.1016/j.ijcchd.2025.100566.
[3] Prajjwal P, Marsool MDM, Yadav V, et al. Neurological, cardiac, musculoskeletal, and renal manifestations of scleroderma along with insights into its genetics, pathophysiology, diagnostic, and therapeutic updates. Health Sci Rep 2024;7(4):e2072. doi:10.1002/hsr2.2072.
[4] Expert Panel on Thoracic Imaging Sirajuddin A, Donnelly EF, et al. ACR Appropriateness Criteria® Suspected Pulmonary Hypertension. J Am Coll Radiol 2017;14(5S):S350–61. doi:10.1016/j.jacr.2017.01.040.
[5] Gonzalez-Hermosillo LM, Cueto-Robledo G, Roldan-Valadez E, et al. Right heart catheterization (RHC): A comprehensive review of provocation tests and hepatic hemodynamics in patients With pulmonary hypertension (PH). Curr Probl Cardiol 2022;47(12):101351. doi:10.1016/j.cpcardiol.2022.101351.
[6] Al-Khazali HM, Christensen RH, Chaudhry BA, et al. Effects of PDE-3 inhibition in persistent post-traumatic headache: evidence of cAMP-dependent signaling. J Headache Pain 2024;25(1):56. doi:10.1186/s10194-024-01762-x.
[7] O'Callaghan DS, Gaine SP. Combination therapy and new types of agents for pulmonary arterial hypertension. Clin Chest Med 2007;28(1):169–85. doi:10.1016/j.ccm.2006.11.011.
[8] Maron BA, Abman SH, Elliott CG, et al. Pulmonary arterial hypertension: Diagnosis, treatment, and novel advances. Am J Respir Crit Care Med 2021;203(12):1472–87. doi:10.1164/rccm.202012-4317SO.
[9] Ribas Sola J, Sánchez-Corral Mena MÁ, Riera-Mestre A. Update in the management of chronic thrombo-embolic pulmonary hypertension. Actualización del abordaje de la hipertensión pulmonar tromboembólica crónica. Med Clin (Barc) 2024;162(3):126–33. doi:10.1016/j.medcli.2023.08.006.
[10] Asif S, Wenhui Y, ur-Rehman S, et al. Advancements and prospects of machine learning in medical diagnostics: unveiling the future of diagnostic precision. Arch Computat Methods Eng 2025;32:853–83. doi:10.1007/s11831-024-10148-w.
[11] Asaduzzaman M, Alom MK, Karim ME. ALZENET: Deep learning-based early prediction of Alzheimer's disease through magnetic resonance imaging analysis. Telemat Inform Rep 2025;17:100189. doi:10.1016/j.teler.2025.100189.
[12] Sharma V, Gupta SK, Shukla KK. Deep learning models for tuberculosis detection and infected region visualization in chest X-ray images. Intell Med 2024;4(2):104–13. doi:10.1016/j.imed.2023.06.001.
[13] Khaled K, Singla M. Predictive analysis of groundwater resources using random forest regression. J Artif Intell Metaheuristics 2025;9:11–19. doi:10.54216/JAIM.090102.

[14] Elshabrawy M. A review on waste management techniques for sustainable energy production. Metaheuristic Optim Rev 2025;3:47–58.
[15] Sharma A, Lim WH, El-Kenawy E-SM, et al. Identification of photovoltaic module parameters by implementing a novel teaching learning based optimization with unique exemplar generation scheme (TLBO-UEGS). Energy Rep 2023;10:1485–506. doi:10.1016/j.egyr.2023.08.019.
[16] Mousavi SMH. Synthetic data generation of body motion data by neural gas network for emotion recognition. arXiv preprint, 2025. doi:10.48550/arXiv.2503.14513.
[17] El-Kenawy ESM, Khodadadi N, Mirjalili S, et al. Greylag goose optimization: Nature-inspired optimization algorithm. Expert Syst Appl 2024;238:122147. doi:10.1016/j.eswa.2023.122147.
[18] Fadilah A, Putri VYS, Puling IMDR, et al. Assessing the precision of machine learning for diagnosing pulmonary arterial hypertension: a systematic review and meta-analysis of diagnostic accuracy studies. Front Cardiovasc Med 2024;11:1422327. doi:10.3389/fcvm.2024.1422327.
[19] Islam H, Iqbal MS, Hossain MM. Blood pressure abnormality detection and Interpretation utilizing explainable artificial intelligence. Intell Med 2025;5(1):54–65. doi:10.1016/j.imed.2024.09.005.
[20] Adeoye J, Su Y. Deep learning with data transformation improves cancer risk prediction in oral precancerous conditions. Intell Med 2025;5(2):141–50. doi:10.1016/j.imed.2024.11.003.
[21] Wang M, Wang J, Hu Y, et al. Detection of pulmonary hypertension with six training strategies based on deep learning technology. Comput Intell 2022;38(5):1684–706. doi:10.1111/coin.12527.
[22] Ge B, Yang H, Ma P, et al. Detection of pulmonary hypertension associated with congenital heart disease based on time-frequency domain and deep learning features. Biomed Signal Process Control 2023;81:104316. doi:10.1016/j.bspc.2022.104316.
[23] Aras MA, Abreau S, Mills H, et al. Electrocardiogram detection of pulmonary hypertension using deep learning. J Card Fail 2023;29(7):1017–28. doi:10.1016/j.cardfail.2022.12.016.
[24] Kıvrak T, Yagmur B, Erken H, et al. Pulmonary hypertension classification using artificial intelligence and chest X-ray: ATA AI STUDY-1. medRxiv preprint, 2023. doi:10.1101/2023.04.14.23288561.
[25] Jimenez-Del-Toro O, Dicente Cid Y, Platon A, et al. A lung graph model for the radiological assessment of chronic thromboembolic pulmonary hypertension in CT. Comput Biol Med 2020;125:103962. doi:10.1016/j.compbiomed.2020.103962.
[26] Nazir A, Naaz R, Qureshi S, Nazir N. HRCT chest analysis for detection of pulmonary arterial hypertension in COVID-19 patients using convolutional neural networks. In: 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT). IEEE; 2022. p. 1–5. doi:10.1109/GCAT55367.2022.9972033.
[27] Lee F, Mielniczuk LM. Pulmonary hypertension due to left heart disease-a practical approach to diagnosis and management. Can J Cardiol 2021;37(4):572–84. doi:10.1016/j.cjca.2020.11.003.
[28] Guazzi M, Borlaug BA. Pulmonary hypertension due to left heart disease. Circulation 2012;126(8):975–90. doi:10.1161/CIRCULATIONAHA.111.085761.
[29] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks. arXiv preprint, 2018. doi:10.48550/arXiv.1801.04381.
[30] Goldberger J, Hinton GE, Roweis S, et al. Neighbourhood components analysis. In: Proceedings of the 18th International Conference on Neural Information Processing Systems (NIPS'04). Cambridge, MA, USA: MIT Press; 2004. p. 513–20.
[31] Liu H, Setiono R. Chi2: Feature selection and discretization of numeric attributes. In: Proceedings of 7th IEEE international conference on tools with artificial intelligence. IEEE; 1995. p. 388–91. doi:10.1109/TAI.1995.479783.
[32] Maillo J, Ramírez S, Triguero I, et al. kNN-IS: An iterative spark-based design of the k-nearest neighbors classifier for big data. Knowl Based Syst 2017;117:3–15. doi:10.1016/j.knosys.2016.06.012.
[33] Vapnik V. The support vector method of function estimation. Nonlinear Model 1998:55–85. doi:10.1007/978-1-4615-5703-6_3.
[34] Vapnik V. The Nature of Statistical Learning Theory. Springer Science & Business Media; 2013. doi:101007/978-1-4757-3264-1_1.
[35] Dogan A, Akay M, Barua PD, et al. PrimePatNet87: Prime pattern and tunable q-factor wavelet transform techniques for automated accurate EEG emotion recognition. Comput Biol Med 2021;138:104867. doi:10.1016/j.compbiomed.2021.104867.
[36] Singh KN, Mantri JK. A clinical decision support system using rough set theory and machine learning for disease prediction. Intell Med 2024;4(3):200–8. doi:10.1016/j.imed.2023.08.002.
[37] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Commun ACM 2017;60(6):84–90. doi:10.1145/306538.
[38] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2021. p. 10012–22.
[39] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint, 2020. doi:10.48550/arXiv.2010.11929.
[40] Liu Z, Mao H, Wu CY, et al. A convnet for the 2020s. In: Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA; 2022. p. 11966–76. doi:10.1109/CVPR52688.2022.01167.
[41] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA; 2016. p. 770–8. doi:10.1109/CVPR.2016.90.