Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Lung Cancer Classification from CT Images: A Comparative Study of SVM, Modified CNN, and EfficientNet–SVM Hybrid Models

Nuran Binte Habib
Chittagong University of Engineering & Technology
Shahida Akther
Chittagong University of Engineering & Technology
Rabiul Islam
rislam@olemiss.edu

University of Mississippi

Article

Keywords: Lung Cancer Detection, Computed Tomography, Convolutional Neural Networks, Support Vector Machine, Class Imbalance, SMOTE, Batch Normalization, Computer-Aided Diagnosis

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11367229/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Lung Cancer Classification from CT Images: A Comparative Study of SVM, Modified CNN, and EfficientNet–SVM Hybrid Models

Nuran Binte Habib[1], Shahida Akther[1], Rabiul Islam[1,2*]

[1*]Department of Electrical and Electronic Engineering, Chittagong University of Engineering and Technology (CUET), Chittagong, 4349, Bangladesh.

[2*]Department of Electrical and Computer Engineering, University of Mississippi, University, Mississippi, 38677, USA.

*Corresponding author(s). E-mail(s): rislam@olemiss.edu;

Abstract

This paper proposes a computer-aided diagnosis (CAD) framework that combines four complementary classification paradigms on the publicly available IQ-OTH/NCCD CT dataset with 1,097 slices across benign, malignant, and normal classes from 110 patients. The framework comprises: (i) an image-processing pipeline employing Gaussian smoothing, histogram equalization, Canny edge detection, and morphological closing, (ii) Otsu-based lung segmentation followed by handcrafted feature extraction combining statistical Local Binary Pattern (LBP), gradient, and shape descriptors, (iii) a Support Vector Machine (SVM) classifier benchmarked under a systematic ablation of three kernels and six class-balancing strategies (SMOTE, ADASYN, Borderline-SMOTE, Random Over-Sampling, SMOTE–Tomek, and cost-sensitive class weighting), (iv) a baseline convolutional neural network (CNN) and a modified CNN architecture enriched with batch normalization, and (v) a hybrid EfficientNet–SVM model that separates deep representation learning from maximum-margin decision-boundary fitting. The modified CNN achieves the highest overall test accuracy of 99.41% with a macro F1-score of 0.99, while the SVM with the RBF kernel under Random Over-Sampling reaches 98.81% balanced accuracy and the hybrid CNN–SVM attains 98.00% test accuracy. The result demonstrates both robustness across classes and competitiveness against recent state-of-the-art models on the same dataset.

Keywords: Lung Cancer Detection, Computed Tomography, Convolutional Neural Networks, Support Vector Machine, Class Imbalance, SMOTE, Batch Normalization, Computer-Aided Diagnosis

1 Introduction

According to the Global Cancer Statistics released by the American Cancer Society in 2024, lung cancer continues to be the leading cause of cancer mortality worldwide, accounting for approximately 2.5 million new cases (one in eight cancers) and 1.8 million deaths (one in five cancer deaths) annually [1]. In Bangladesh, cancer is responsible for nearly 10% of all deaths and ranks as the sixth most frequent cause of mortality among non-communicable diseases [2]. Among Bangladeshi males, lung cancer constitutes 11.1% of all newly diagnosed cancers, making it the second most prevalent malignancy; in females, it ranks sixth with 4.6% of new cases [3].

Early detection is decisive for prognosis. The five-year survival rate falls precipitously from approximately 65% when the disease is identified at a localised stage to roughly 9% once it has metastasised [4]. Pulmonary nodule analysis is therefore the cornerstone of screening protocols [5]. Clinically, lung cancer is broadly categorised into non-small-cell lung cancer (NSCLC) and small-cell lung cancer (SCLC), with NSCLC accounting for the majority of cases and typically progressing more gradually than the aggressive SCLC subtype. Diagnostic modalities include physical examination, chest X-ray [6], magnetic resonance imaging [7], low-dose computed tomography (CT) [8], bronchoscopy, tissue biopsy with histopathological confirmation, and molecular testing for actionable biomarkers [9]. Among these, CT remains the modality of choice for early detection and detailed lung-nodule characterisation [10].

Despite the diagnostic value of CT, two clinical realities motivate the development of automated decision-support tools. First, the volumetric nature of CT acquisitions (typically 80–200 slices per case) imposes a heavy cognitive load on radiologists. Second, subtle early-stage nodules are easily overlooked or confused with benign lesions, contributing to inter-observer variability. Machine learning, and particularly deep learning, has emerged as an effective avenue for mitigating these limitations by enabling reproducible, high-throughput, and quantitative image analysis.

Recent studies on the IQ-OTH/NCCD dataset report consistently high accuracies, but three methodological gaps continue to limit the comparability of these results. First, most reports evaluate either a deep learning model in isolation [11–14] or a classical machine learning pipeline alone [15, 16]; comparatively few works [17] provide a unified comparison spanning handcrafted feature pipelines, end-to-end CNNs, and hybrid CNN–SVM architectures on a single benchmark. Second, the choice of class-imbalance handling is rarely justified through a systematic ablation study; instead, a single technique (often plain SMOTE) is adopted without comparing any alternative ones. Third, the dominant evaluation reporting is restricted to accuracy and macro-averaged F1-score, with limited use of imbalance-aware metrics such as the Matthews Correlation Coefficient (MCC) and Cohen's κ that more faithfully characterise multi-class performance under skewed distributions [18, 19].

This work addresses the above gaps and contributes the following.

1. A carefully engineered image-processing pipeline (Gaussian filtering, histogram equalisation, Canny edge detection, morphological closing) is combined with Otsu thresholding-based lung segmentation and a four-category handcrafted feature set (statistical, Local Binary Patterns, gradient, and shape descriptors). The pipeline is designed to be deterministic, computationally lightweight, and reproducible.
2. A systematic 6×3 ablation grid is reported on SVM classifiers, covering six imbalance-handling strategies (SMOTE, ADASYN, Borderline-SMOTE, Random Over-Sampling, SMOTE–Tomek, and cost-sensitive class weighting) in combination with three SVM kernels (linear, RBF, polynomial). To the best of our knowledge, this is the most comprehensive imbalance–kernel comparison reported on IQ-OTH/NCCD to date.
3. A Modified CNN architecture enriched with batch normalisation [20] after each convolutional block is shown to attain 99.41% test accuracy without any GAN-based augmentation, outperforming both the baseline CNN of this study and several recent IQ-OTH/NCCD benchmarks.
4. A hybrid EfficientNet-B0–SVM model is presented, where the EfficientNet backbone [21] produces a 1,280-dimensional embedding fed to a kernel SVM. The hybrid model achieves 98.00% test accuracy with perfect recall on malignant cases, exemplifying how representation learning and maximum-margin classification can be decoupled.
5. Beyond standard accuracy, precision, recall, and F1-score, balanced accuracy, macro-averaged AUC-ROC, Matthews Correlation Coefficient (MCC), and Cohen's κ are reported throughout to strengthen evaluation rigour for clinical translation.
6. A unified algorithm (Algorithm 1) consolidating preprocessing, segmentation, feature extraction, imbalance handling, model selection, and evaluation is presented to facilitate reproducibility and adoption.

The remainder of the paper is structured as follows. Section 2 reviews related work, with emphasis on SVM-, CNN-, and hybrid-based approaches to lung-cancer detection. Section 3 details the dataset, preprocessing, segmentation, feature-extraction, class-imbalance, and classification methodologies, and consolidates the framework into Algorithm 1. Section 4 reports experimental results, including the 6×3 ablation grid, the Modified CNN, and the hybrid model. Section **??** discusses findings, compares against recent state-of-the-art works on the same dataset, and articulates limitations. Section 5 concludes with prospects for future research.

2 Related Work

Artificial intelligence has been increasingly integrated into medical imaging in recent years to enhance the precision and efficiency of lung-cancer detection. Early sputum-cytology and chest-X-ray approaches paved the way for low-dose CT (LDCT) screening protocols in the early 2000s. Subsequently, classical machine-learning techniques — including support vector machines (SVM) [15], logistic regression [22], k-nearest neighbours [23], and random forests [24] introduced data-driven nodule classification.

The introduction of deep convolutional architectures [25], recurrent attention networks [26], and fully convolutional networks [10] in the mid-2010s further accelerated progress by enabling automated hierarchical feature extraction. More recently, hybrid pipelines that combine deep features with classical classifiers (e.g. SVM, random forest, XGBoost) have attracted increasing interest [27]. The remainder of this section organises the literature along three axes: SVM-based approaches, CNN-based approaches, and hybrid (modified-CNN) approaches.

2.1 SVM-based Approaches

SVMs have long been a cornerstone of computer-aided diagnosis because of their maximum-margin formulation and their robustness in high-dimensional feature spaces. Kareem et al. [15] evaluated an SVM classifier on the IQ-OTH/NCCD dataset after applying bit-plane slicing, Gaussian filtering, and lung-border detection. They reported 89.88% accuracy, 98.55% precision, 97.84% F1-score, 97.14% sensitivity, and 97.50% specificity, establishing an early performance reference on this dataset though the pipeline depends on manual bit-plane slicing and yields lower accuracy than several later studies on the same dataset. Hussain et al. [16] combined Gray-Level Co-occurrence Matrix (GLCM) texture features with SVMs employing RBF and polynomial kernels on MRI lung images from the Lung Cancer Alliance dataset; flipping, rotation, gamma correction, and histogram equalisation were applied for augmentation; however, no standard classification metrics were reported, and validation was performed on MRI rather than CT, limiting comparability to the present work. Hassan et al. [28] compared SVM, XGBoost, ANN, and logistic regression on IQ-OTH/NCCD with median and Gaussian filtering plus brightness, contrast, flipping, and rotation augmentations, observing that SVM attained the highest accuracy of 99.30%, with precision 99.10% and recall 99.19% although the comparison is missing imbalance-aware metrics such as MCC or Cohen's κ, leaving the models' behaviour on the minority benign class unclear. Collectively, these studies demonstrate the resilience of SVM-based pipelines when paired with discriminative handcrafted features, but typically address only a single class-imbalance strategy.

2.2 CNN-based Approaches

Convolutional neural networks have become the workhorse of automated lung-cancer detection. Al-Yasriy et al. [11] applied AlexNet to IQ-OTH/NCCD with a preprocessing pipeline that included resampling, segmentation, normalisation, and conversion to Hounsfield units, reporting 93.54% accuracy, 95.71% sensitivity, 97.10% precision, and 96.40% F1-score; however, the shallow AlexNet backbone constrains representational capacity relative to other deeper neural architectures. Tasnim et al. [12] designed a custom CNN model that, after SMOTE augmentation and a range of geometric transformations, outperformed ResNet50 and InceptionV3 with validation accuracy of 98% and F1-score of 97%; however, only a single imbalance-handling technique (SMOTE) was evaluated without a comparative ablation across different strategies. Ruprah et al. [13] adopted transfer learning with VGG-16 on an augmented subset of IQ-OTH

/ NCCD, achieving 99% accuracy, but performance leans heavily on ImageNet-pretrained weights, leaving how well the features transfer from natural images to CT. Wadekar and Singh [29] introduced a modified VGG-19 with residual connections trained on the LC25000 histopathological collection, recording 97.73% accuracy and macro precision/recall/F1 of 0.97. Most recently, Shariff et al. [14] demonstrated a CNN augmented with differential augmentation (DA) for non-small-cell lung cancer, reporting 98.78% accuracy though the differential-augmentation procedure adds non-trivial computational overhead during training. Recent works have implemented the state-of-the-art attention- or transformer-augmented CNNs. Lung-EffNet [30] leveraged EfficientNet-B1 for 99.10% accuracy with GAN-based augmentation,and LCxNet [31] integrated NAdam-optimised CNNs with Grad-CAM and t-SNE interpretability, achieving 99.39% accuracy. However, both model rely on GAN-based augmentation or large attention backbones, adding training complexity that the present Modified CNN avoids.

2.3 Hybrid and Modified-CNN Approaches

Hybrid models attempt to exploit the complementary strengths of deep feature extractors and discriminative classifiers. Pawar et al. [27] proposed a hybrid decision tree–CNN system on the LIDC-IDRI dataset, combining Gaussian filtering, Laplacian edge detection, Otsu thresholding, an autoencoder for representation, and a fusion of GLCM and HOG descriptors. Hasan et al. [32] integrated CNN-derived features with a random forest on IQ-OTH/NCCD, achieving 97.10% accuracy through k-fold cross-validation and augmentation. Amouzad [33] explored a benchmark example wherein EfficientNetV2-S features fed an RBF-SVM on the CIFAR-10 dataset, reaching 97.89% accuracy and 0.9789 F1-score; while not lung-specific, this study provided a template for the EfficientNet–SVM hybrid we adopt here. Saikia et al. [17] also reported closely relevant study by coupling VGG-derived features with SVM and Random Forest for four-class lung-nodule classification, attaining 98.70% accuracy. The present work extends this line by replacing VGG with a compound-scaled EfficientNet-B0 backbone and by additionally benchmarking the hybrid against a Modified CNN trained end-to-end.

3 Methods

The proposed framework integrates six sequential stages: (i) dataset acquisition and stratification, (ii) image preprocessing, (iii) lung-region segmentation, (iv) feature extraction, (v) class-imbalance handling and classification by three model families (SVM, CNN, Modified CNN, hybrid CNN–SVM), and (vi) model evaluation. Figure 1 illustrates the overall pipeline, and Algorithm 1 formalises the unified procedure.

3.1 Dataset Description

The Iraq-Oncology Teaching Hospital / National Center for Cancer Diseases (IQ-OTH/NCCD) dataset was compiled over a three-month period in 2019 [11]. It comprises 110 cases distributed across three diagnostic categories: *benign* (15 cases),

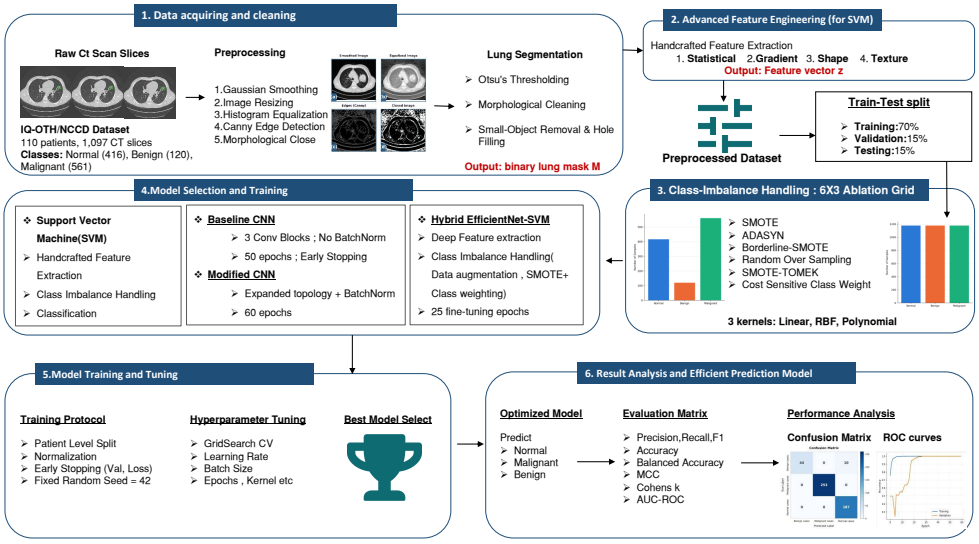


Fig. 1 Workflow Diagram of proposed methodology.

malignant (40 cases), and *normal* (55 cases). Each case contributes between 80 and 200 axial CT slices, captured on a Siemens SOMATOM scanner with the following acquisition protocol: tube voltage 120 kVp, slice thickness 1 mm, and window settings of 350–1,200 HU width and 50–600 HU centre,note that the 80–200 figure refers to the raw acquired CT volume per case; for annotation and experimentation, a curated subset of diagnostically representative slices per patient was selected, yielding the 1,097 total analysed slices used throughout this study.. Images are stored in DICOM format and de-identified at source. Patient demographics include age, gender, residence (primarily Baghdad, Wasit, Diyala, Salahuddin, and Babylon), and occupation. In total, the dataset contains 1,097 annotated slices distributed as 416 normal, 120 benign, and 561 malignant images, yielding an imbalance ratio of approximately 4.67:1 between the majority (malignant) and minority (benign) classes. Although some recent reports cite 1,190 images, the arithmetic of the per-class counts confirms 1,097 as the correct total, and this number is adopted throughout the paper. Figure 2 shows representative samples for each class.

3.2 Image Preprocessing

To suppress acquisition noise and enhance lesion conspicuity, every CT slice I is processed through the following sequence.

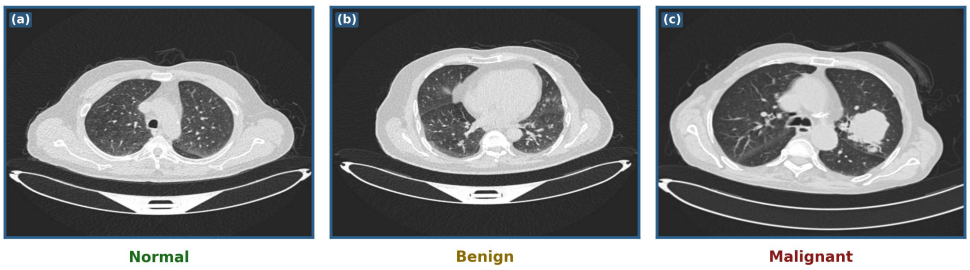


Fig. 2 Representative axial CT slices from the IQ-OTH/NCCD dataset: (a) normal lung, (b) benign nodule, (c) malignant lesion.

Gaussian smoothing

A two-dimensional isotropic Gaussian filter G_σ with $\sigma = 1$ is convolved with the input:

$$I_{\text{sm}}(x, y) = (G_\sigma * I)(x, y) \qquad G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \tag{1}$$

Image resizing

All slices are resampled to 256×256 pixels via bilinear interpolation to ensure dimensional uniformity across the dataset.

Histogram equalisation

Let $h(i)$ denote the histogram of intensity $i \in \{0, \ldots, L-1\}$. The cumulative distribution $H(i) = \sum_{k=0}^{i} h(k)/N$ is used to construct the transformation $T(i) = \lfloor (L-1)\, H(i) \rfloor$, yielding an equalised image $I_{\text{eq}}(x, y) = T(I_{\text{sm}}(x, y))$ with redistributed intensities and improved global contrast.

Canny edge detection

Edges are extracted by hysteresis thresholding on the gradient magnitude of the equalised image, with low and high thresholds of 50 and 150 respectively.

Morphological closing

A binary closing operation (dilation followed by erosion) with an elliptical structuring element B is applied to the edge map:

$$I_{\text{close}} = (I_{\text{eq}} \oplus B) \ominus B \tag{2}$$

where $\oplus$ and $\ominus$ denote dilation and erosion respectively. The closing step bridges discontinuities along nodule boundaries and improves the integrity of subsequently segmented regions.

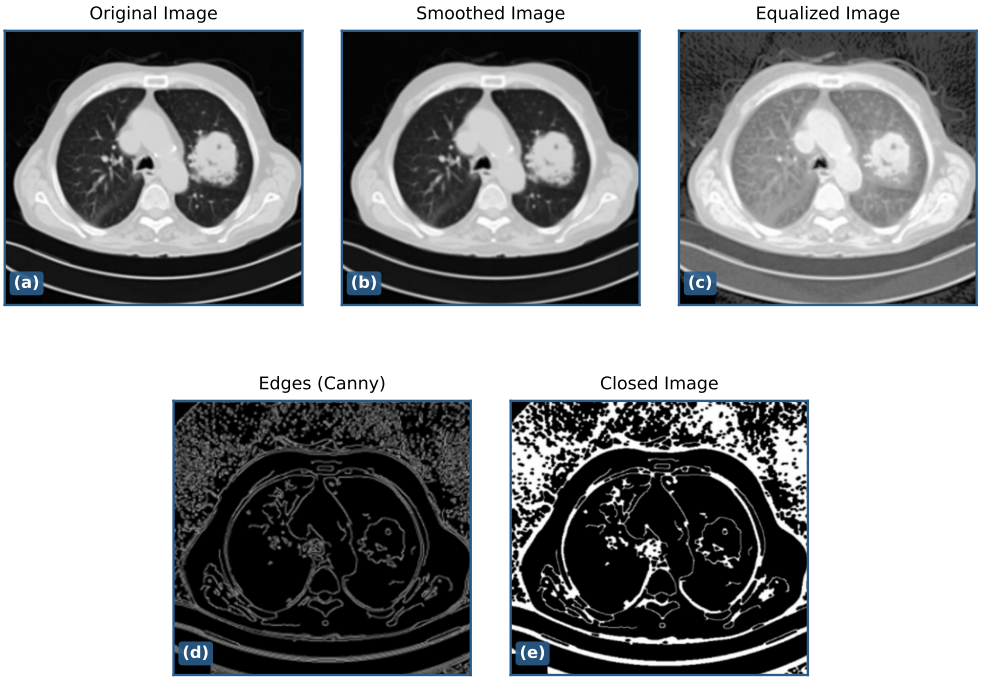


Fig. 3 Image preprocessing stages: (a) original CT slice, (b) Gaussian-smoothed, (c) histogram-equalised, (d) Canny edges, (e) morphologically closed.

3.3 Lung-Region Segmentation

The segmentation stage isolates the parenchyma to focus subsequent analysis on the diagnostically relevant region. Three complementary procedures are applied.

Otsu's thresholding

Otsu's method [34] selects the threshold $t^\star$ that maximises the between-class variance of the bimodal intensity histogram. With class probabilities $\omega_0(t) = \sum_{i=0}^{t} p_i$ and $\omega_1(t) = 1 - \omega_0(t)$, and class means $\mu_0(t)$ and $\mu_1(t)$,

$$t^\star = \arg\max_{0 \le t \le L-1} \omega_0(t)\,\omega_1(t)\,\big(\mu_0(t) - \mu_1(t)\big)^2 \tag{3}$$

Morphological cleaning

Opening and closing operations smooth boundaries and remove spurious foreground pixels arising from imaging artefacts.

Small-object removal and hole filling

Objects with area below 500 pixels are discarded, and holes smaller than 1,000 pixels are filled, producing a final binary mask $M \in \{0, 255\}^{H \times W}$ that delineates the lung region for downstream feature extraction.

3.4 Feature Extraction

Four families of features are extracted from each segmented region to yield a feature vector $\mathbf{z} \in \mathbb{R}^d$.

3.4.1 Statistical features

Let $\{I_n\}_{n=1}^{N}$ denote the intensities within the segmented region of size N. The first seven statistical descriptors are

$$\mu = \frac{1}{N}\sum_{n=1}^{N} I_n \qquad \sigma = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(I_n - \mu)^2} \tag{4}$$

$$\sigma^2 = \frac{1}{N}\sum_{n=1}^{N}(I_n - \mu)^2 \qquad \widetilde{I} = \operatorname{median}\{I_1, \ldots, I_N\} \tag{5}$$

$$\min = \min_n I_n \qquad \max = \max_n I_n \tag{6}$$

together with selected percentiles P_q ($q \in \{25, 50, 75\}$), defined as the smallest value below which $q\%$ of intensities fall. Higher-order moments (skewness γ_1 and excess kurtosis γ_2) are also computed to capture asymmetry and tail-heaviness:

$$\gamma_1 = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{I_n - \mu}{\sigma}\right)^3 \quad \gamma_2 = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{I_n - \mu}{\sigma}\right)^4 - 3 \tag{7}$$

3.4.2 Local Binary Patterns (LBP)

The classical LBP operator of Ojala et al. [35] encodes local texture by thresholding P circularly arranged neighbours g_p at radius R against a central pixel g_c:

$$\mathrm{LBP}_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c)\, 2^p \qquad s(z) = \begin{cases} 1, & z \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

To improve rotation invariance, the uniform rotation-invariant variant $\mathrm{LBP}^{riu2}_{P,R}$ is also adopted:

$$\mathrm{LBP}^{riu2}_{P,R} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & \text{if } U(\mathrm{LBP}_{P,R}) \leq 2 \\ P + 1 & \text{otherwise} \end{cases} \tag{9}$$

where $U(\cdot)$ counts the number of 0/1 transitions in the circular bit-string.

3.4.3 Gradient features

First-order spatial derivatives are estimated with Sobel kernels, producing

$$\nabla_x I = \frac{\partial I}{\partial x}, \quad \nabla_y I = \frac{\partial I}{\partial y}, \quad |\nabla I| = \sqrt{(\nabla_x I)^2 + (\nabla_y I)^2}, \quad \theta = \arctan_2(\nabla_y I, \nabla_x I) \tag{10}$$

3.4.4 Shape features

Six descriptors are computed from the largest connected component of the binary mask: area A, perimeter P, compactness $C = 4\pi A/P^2$, aspect ratio $AR = w_{BB}/h_{BB}$ (using the bounding-box width and height), extent $E = A/(w_{BB} \cdot h_{BB})$, and solidity $S = A/A_{\text{convex}}$, where A_{convex} is the area of the convex hull. These descriptors characterise the geometry of the lung region and are particularly informative for distinguishing nodular from diffuse pathology.

3.5 Class-Imbalance Handling

The dataset exhibits a substantial imbalance, with benign cases representing only 10.9% of slices (Table 1). Six strategies are evaluated.

SMOTE

For each minority-class sample $\mathbf{x}_i$, a k-nearest neighbour $\mathbf{x}_{nn}$ within the minority class is selected, and a synthetic sample can be generated using the equation below. [36]

$$\mathbf{x}_{\text{new}} = \mathbf{x}_i + \lambda\big(\mathbf{x}_{nn} - \mathbf{x}_i\big) \qquad \lambda \sim \mathcal{U}[0, 1] \tag{11}$$

ADASYN

A density-dependent weight $\hat{r}_i$ is assigned to each minority sample $\mathbf{x}_i$ according to the proportion of majority-class neighbours Δ_i among its k nearest neighbours [37],

$$r_i = \frac{\Delta_i}{k}, \qquad \hat{r}_i = \frac{r_i}{\sum_j r_j}, \qquad g_i = \lceil \hat{r}_i \cdot G \rceil \tag{12}$$

with G the total number of synthetic samples to be generated. ADASYN thereby focuses augmentation on harder-to-learn minority regions.

Borderline-SMOTE

Synthetic samples are generated only from *danger* samples, defined as minority instances for which the majority of k nearest neighbours belong to the majority class, sharpening the decision boundary [38].

Random Over-Sampling (ROS)

Minority samples are duplicated uniformly at random until the desired class ratio is achieved. While conceptually simple, ROS is more prone to overfitting than synthetic methods.

SMOTE–Tomek

SMOTE oversampling is combined with the removal of Tomek links pairs $(\mathbf{x}_i, \mathbf{x}_j)$ from opposite classes that are each other's nearest neighbour so as to denoise the resulting feature space[39].

Cost-sensitive class weighting

Rather than resampling, the SVM loss is reweighted with class-frequency-inverse weights $w_k = N/(K \cdot N_k)$, where K is the number of classes and N_k the number of samples in class k.

3.6 Classification Models

3.6.1 Support Vector Machine

The SVM seeks an optimal hyperplane $\mathbf{w}^\top \mathbf{x} + b = 0$ that maximises the margin between classes [40]. With soft-margin slacks $\xi_i \geq 0$ and regularisation parameter C, the primal becomes

$$\min_{\mathbf{w}, b, \boldsymbol{\xi}} \ \tfrac{1}{2}\|\mathbf{w}\|^2 + C \sum_{i=1}^{N} \xi_i \quad \text{s.t.} \quad y_i(\mathbf{w}^\top \Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \tag{13}$$

with $\Phi(\cdot)$ a kernel-induced feature map. Three kernels are evaluated:

$$K_{\text{lin}}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^\top \mathbf{x}_j \tag{14}$$

$$K_{\text{poly}}(\mathbf{x}_i, \mathbf{x}_j) = \big(\gamma\, \mathbf{x}_i^\top \mathbf{x}_j + r\big)^d \tag{15}$$

$$K_{\text{rbf}}(\mathbf{x}_i, \mathbf{x}_j) = \exp\!\big(-\gamma\, \|\mathbf{x}_i - \mathbf{x}_j\|^2\big) \tag{16}$$

Hyperparameters $C \in \{0.1, 1, 10\}$, $\gamma \in \{\text{scale}, 0.01, 0.1\}$, and polynomial degree $d \in \{2, 3\}$ are tuned by stratified 5-fold cross-validation with balanced accuracy as the selection criterion.

3.6.2 Baseline CNN

The baseline network comprises three convolutional blocks (32, 64, and 128 filters with 3×3 kernels), each followed by 2×2 max-pooling, a flatten layer, a dense layer of 256 units with ReLU activation, dropout with rate 0.2, a 128-unit dense layer, a second dropout, and a softmax output of three units. For a single convolutional neuron with weights $\mathbf{W}$, receptive field $\mathbf{X}$, and bias b, the operation reduces to

$$Z = \sum_i W_i X_i + b, \qquad a = \text{ReLU}(Z) = \max(0, Z) \tag{17}$$

with max-pooling $P = \max(\mathbf{X})$ over the pooling window. The output layer employs the softmax

$$\text{softmax}(z_j) = \frac{\exp(z_j)}{\sum_{k=1}^{K} \exp(z_k)} \tag{18}$$

and the network is trained by minimising the categorical cross-entropy loss

$$\mathcal{L}_{\mathrm{CE}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{ik}\log p_{ik} \tag{19}$$

3.6.3 Modified CNN with Batch Normalization

The Modified CNN replicates the baseline topology but inserts a batch-normalisation layer [20] after every convolutional operation, prior to ReLU activation. For a mini-batch $\mathcal{B} = \{x_1, \ldots, x_m\}$, batch normalisation computes

$$\mu_{\mathcal{B}} = \tfrac{1}{m}\sum_{i=1}^{m} x_i, \qquad \sigma^2_{\mathcal{B}} = \tfrac{1}{m}\sum_{i=1}^{m}(x_i - \mu_{\mathcal{B}})^2 \tag{20}$$

$$\hat{x}_i = \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma^2_{\mathcal{B}} + \varepsilon}}, \qquad y_i = \gamma\,\hat{x}_i + \beta \tag{21}$$

with learnable scale γ and shift β. By reducing internal covariate shift, batch normalisation accelerates convergence, enables larger learning rates, and acts as an implicit regulariser [20]. Within our experiments, this single architectural modification translates into a 0.62 percentage-point improvement in test accuracy (from 98.79% to 99.41%) while reducing training time per epoch.

3.6.4 Hybrid CNN–SVM Model

The hybrid model employs EfficientNet-B0 [21], pre-trained on ImageNet, as a frozen feature extractor. Let $f_\theta : \mathbb{R}^{H\times W\times 3} \to \mathbb{R}^{1280}$ denote the backbone followed by global average pooling. For each input $\mathbf{x}_i$, the embedding $\mathbf{z}_i = f_\theta(\mathbf{x}_i)$ is computed and passed to an SVM trained on $\{(\mathbf{z}_i, y_i)\}$. The dual SVM problem is

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j\, y_i y_j\, K(\mathbf{z}_i, \mathbf{z}_j), \qquad 0 \le \alpha_i \le C, \quad \sum_i \alpha_i y_i = 0 \tag{22}$$

with prediction $\hat{y}(\mathbf{z}) = \operatorname{sign}\big(\sum_i \alpha_i y_i K(\mathbf{z}_i, \mathbf{z}) + b\big)$. For the three-class IQ-OTH/NCCD problem, three one-vs-one binary SVMs are trained and voting determines the final label. SMOTE-based balancing and class weighting are combined with grid-search hyperparameter tuning over $C \in \{0.1, 1, 10, 100\}$, $\gamma \in \{\text{scale}, \text{auto}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$, and kernel $\in \{\text{RBF}, \text{linear}\}$, with stratified 5-fold cross-validation and macro F1-score as the selection criterion.

3.7 Evaluation Metrics

For a multi-class confusion matrix with entries c_{ij} (true class i, predicted class j), per-class counts are $\mathrm{TP}_k = c_{kk}$, $\mathrm{FP}_k = \sum_{i\neq k} c_{ik}$, and $\mathrm{FN}_k = \sum_{j\neq k} c_{kj}$. The principal

metrics are

$$\text{Accuracy} = \frac{\sum_k \text{TP}_k}{N} \tag{23}$$

$$\text{Precision}_k = \frac{\text{TP}_k}{\text{TP}_k + \text{FP}_k} \quad \text{Recall}_k = \frac{\text{TP}_k}{\text{TP}_k + \text{FN}_k} \tag{24}$$

$$F1_k = \frac{2\,\text{Precision}_k\,\text{Recall}_k}{\text{Precision}_k + \text{Recall}_k} \tag{25}$$

$$\text{Balanced Accuracy} = \frac{1}{K}\sum_{k=1}^{K} \text{Recall}_k \tag{26}$$

To strengthen the evaluation within class imbalance, two correlation-based metrics are additionally reported. Cohen's κ [41] corrects accuracy for chance agreement:

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \qquad p_o = \text{Accuracy}, \quad p_e = \frac{1}{N^2}\sum_{k=1}^{K}\Big(\sum_i c_{ik}\Big)\Big(\sum_j c_{kj}\Big). \tag{27}$$

The multi-class Matthews Correlation Coefficient [19], defined as

$$\text{MCC} = \frac{\sum_k c_{kk}\, N - \sum_{k,l} c_{kl}\, c_{lk}}{\sqrt{\big(N^2 - \sum_k (\sum_l c_{kl})^2\big)\big(N^2 - \sum_k (\sum_l c_{lk})^2\big)}}, \tag{28}$$

provides a single scalar in $[-1, 1]$ that is more informative than F1 when classes are highly imbalanced [18].

3.8 Unified Algorithm

To facilitate reproducibility, Algorithm 1 consolidates the preprocessing, segmentation, feature-extraction, imbalance-handling, training, and evaluation steps of the four classification paradigms into a single procedure.

4 Results

4.1 Experimental Setup

All experiments are conducted on the IQ-OTH/NCCD dataset under a stratified train/validation/test split of 70/15/15. The classical pipeline is implemented in Python using scikit-learn 1.3, with image processing in OpenCV 4.8 and `scikit-image` 0.21. The deep-learning pipeline is built upon TensorFlow 2.13 with the Keras API. All training is performed on a workstation equipped with an NVIDIA RTX-class GPU. For the SVM ablation, hyperparameters are tuned by 5-fold stratified cross-validation; for the CNN models, training is run for 50–60 epochs with the Adam optimiser at an initial learning rate of 10^{-3}, early stopping with a patience of 5 epochs, and a batch size of 32. Random seeds are fixed (`random_state=42`) for reproducibility.

Algorithm 1 Unified Framework for Lung Cancer Classification.

Require: Dataset $\mathcal{D} = \{(I_i, y_i)\}_{i=1}^{N}$ with $y_i \in \{\text{benign}, \text{malignant}, \text{normal}\}$
Require: Set of imbalance strategies $\mathcal{T} = \{\text{SMOTE, ADASYN, B-SMOTE, ROS, S-Tomek, class weights}\}$
Require: Set of SVM kernels $\mathcal{K} = \{\text{linear}, \text{RBF}, \text{poly}\}$
Ensure: Trained model $M^\star$ and its predictions $\hat{\mathbf{y}}$ on test set

Phase 1: Preprocessing
for each image $I_i \in \mathcal{D}$ **do**
  $I_i^{(1)} \leftarrow G_\sigma * I_i$ ▷ Gaussian smoothing
  $I_i^{(2)} \leftarrow \text{Resize}(I_i^{(1)}, 256 \times 256)$
  $I_i^{(3)} \leftarrow \text{HistEq}(I_i^{(2)})$ ▷ Histogram equalisation
  $I_i^{(4)} \leftarrow \text{Canny}(I_i^{(3)}, t_{\text{lo}}{=}50, t_{\text{hi}}{=}150)$
  $I_i^{(5)} \leftarrow (I_i^{(4)} \oplus B) \ominus B$ ▷ Morphological closing
end for
Phase 2: Segmentation
for each $I_i^{(5)}$ **do**
  $t_i^\star \leftarrow \arg\max_t \, \omega_0(t)\omega_1(t)(\mu_0(t) - \mu_1(t))^2$
  $M_i \leftarrow \text{BinariseAndClean}(I_i^{(5)}, t_i^\star)$
end for
Phase 3: Feature extraction (handcrafted pipeline)
for each M_i **do**
  $\mathbf{z}_i \leftarrow [\text{Statistical}, \text{LBP}_{P,R}^{riu2}, \nabla I, \text{Shape}]$
end for
Phase 4: Classical pipeline (6×3 ablation grid)
for each $\tau \in \mathcal{T}$ and each $K \in \mathcal{K}$ **do**
  $(\mathbf{Z}', \mathbf{y}') \leftarrow \tau(\mathbf{Z}, \mathbf{y})$
  Split $(\mathbf{Z}', \mathbf{y}')$ into train (70%) / test (30%) stratified
  $\text{SVM}_{\tau,K} \leftarrow \text{GridSearchCV}(C, \gamma, d; \text{5-fold}, \text{balanced acc.})$
end for
Phase 5: Deep-learning pipelines
Train baseline CNN on raw images with $\mathcal{L}_{\text{CE}}$
Train Modified CNN (BN after each conv block) with $\mathcal{L}_{\text{CE}}$
Compute embeddings $\mathbf{e}_i \leftarrow \text{GlobalAvgPool}(f_{\text{EffNet-B0}}(I_i))$
Train SVM on $\{(\mathbf{e}_i, y_i)\}$ with grid-search and 5-fold CV
Phase 6: Evaluation and selection
for each candidate model M **do**
  Compute Accuracy, Macro-F1, Balanced Acc., MCC, κ, macro AUC-ROC
end for
$M^\star \leftarrow \arg\max_M \text{Macro-F1}_{\text{val}}(M)$
return $M^\star$ and $\hat{\mathbf{y}}_{\text{test}} \leftarrow M^\star(\mathbf{X}_{\text{test}})$

Table 1 Class distribution before and after each sampling technique (number of cases).

Class	Original	SMOTE	ADASYN	B-SMOTE	ROS	SMOTE–Tomek
Normal	416	561	562	561	561	561
Benign	120	561	580	561	561	561
Malignant	561	561	561	561	561	561

4.2 Preprocessing and Segmentation Outputs

Figure 3 illustrates the sequential effect of preprocessing on a representative slice. Gaussian smoothing visibly suppresses high-frequency noise without erasing dominant edges; histogram equalisation amplifies contrast in the perihilar region; Canny edges delineate the parenchymal contour; and morphological closing eliminates small discontinuities, yielding a clean mask. Otsu thresholding combined with small-object removal and hole filling consistently isolates the lung region across all slices.

4.3 SVM Pipeline: 6×3 Ablation Grid

After applying each resampling strategy to the handcrafted feature set, the per-class case counts (Table 1) confirm that all synthetic methods restore approximate equality, while class weighting leaves the underlying counts untouched and instead alters the loss.

Figure 6 and Figure 7 present the balanced accuracy and macro F1-score, respectively, across the full 6×3 ablation grid, and Figure 8 reports macro-averaged AUC-ROC.

Three regularities emerge consistently:

- The RBF kernel dominates the linear and polynomial kernels for every imbalance strategy, in line with the high-dimensional non-linear structure of the feature space.
- ADASYN, Random Over-Sampling, and Borderline-SMOTE achieve the strongest balanced accuracies, with values exceeding 98% on the RBF kernel.
- Cost-sensitive class weighting yields the weakest performance on this dataset, suggesting that explicit oversampling is necessary when the minority class is severely under-represented.

The scatter plot in Figure 9 visualises the joint behaviour of balanced accuracy and AUC-ROC across all 18 configurations; the upper-right cluster of (ROS, RBF) and (ADASYN, RBF) confirms their joint dominance, while class-weighting alone occupies the lower-left region.

A condensed summary of the ten best-performing configurations is reported in Table 2. The Random-Over-Sampler/RBF configuration attains the highest balanced accuracy of 98.81% and macro F1 of 98.81%, with weighted and macro AUC-ROC values reaching 99.99%. The corresponding per-class metrics are listed in Table 3, and the confusion matrix for this best model is shown in Figure 10.

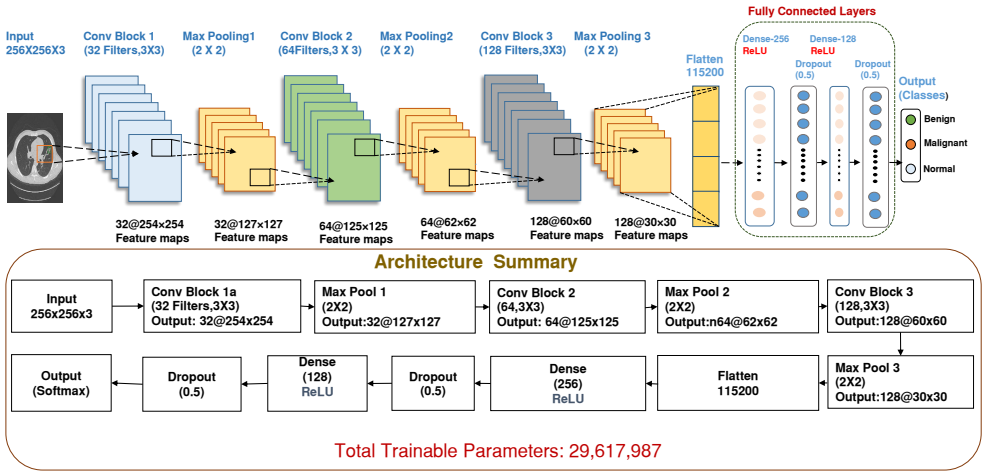


Fig. 4 Architecture of the baseline CNN, comprising three convolution–max-pooling blocks followed by a two-layer fully connected classifier with dropout regularisation.

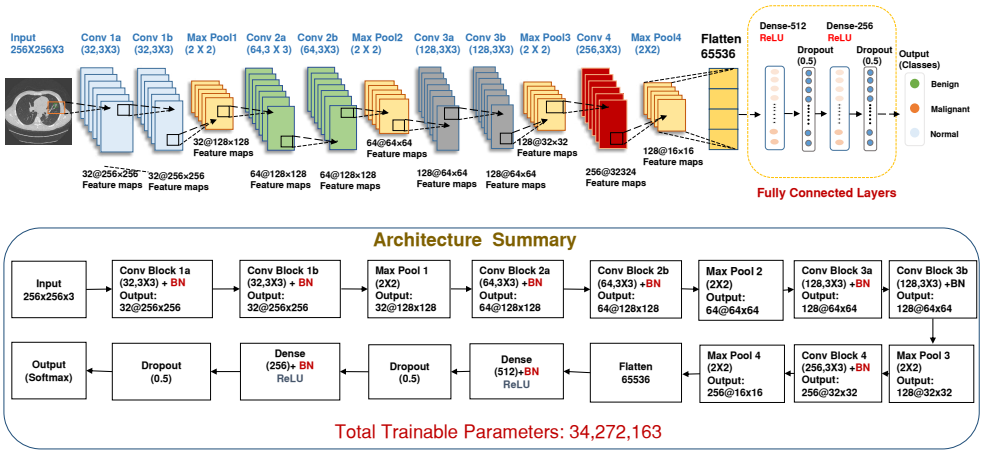


Fig. 5 Architecture of the Modified CNN, incorporating batch normalisation after every convolutional and dense layer to accelerate convergence and improve generalisation.

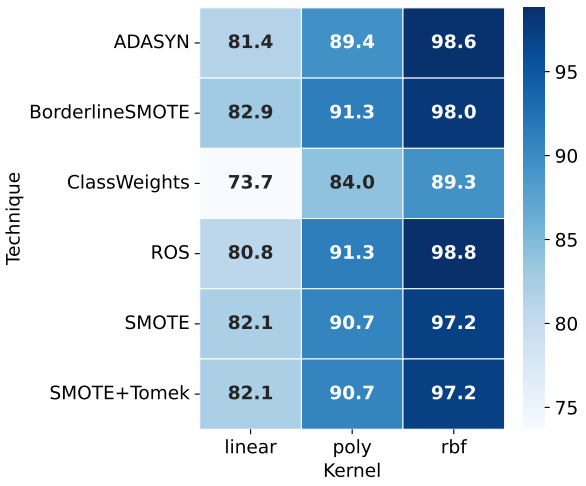


Fig. 6 Balanced accuracy of the SVM classifier across six imbalance-handling strategies and three kernels.

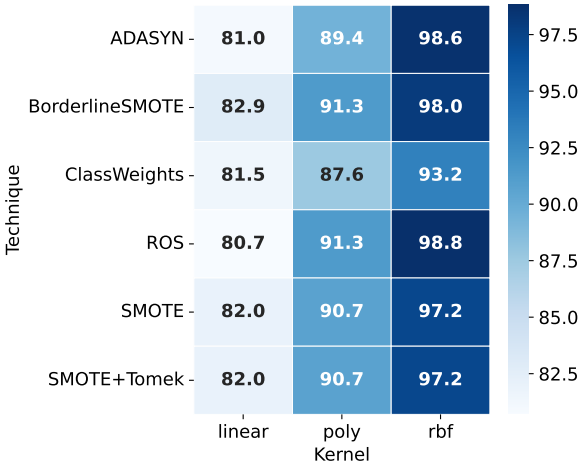


Fig. 7 Macro F1-score of the SVM classifier across the 6×3 ablation grid.

4.4 Baseline CNN

The baseline CNN is trained for 50 epochs with 70% of the data for training, 15% for validation, and 15% for testing. The model (Figure 4) achieves an overall test accuracy of 98.79% with cross-entropy loss of 0.0338. Figure 11 shows the convergence behaviour: both training and validation accuracy rise sharply within the first ten epochs and stabilise thereafter, with no apparent divergence indicative of overfitting.

Per-class metrics in Table 4 reveal perfect classification on malignant cases and slightly reduced performance on benign cases, an unsurprising consequence of the minority-class size. The corresponding confusion matrix is presented in Figure 12.

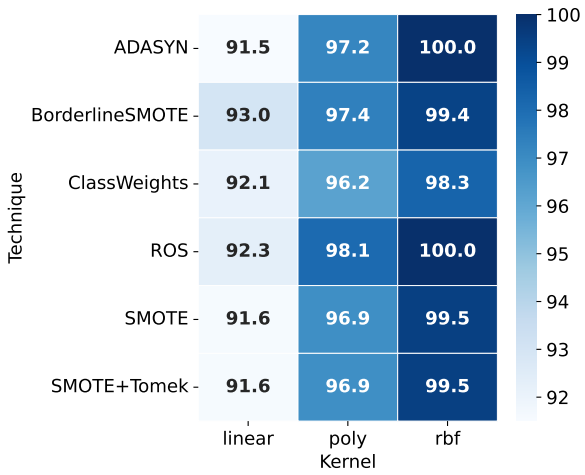


Fig. 8 Heat-map of macro-averaged AUC-ROC across six imbalance strategies and three SVM kernels.

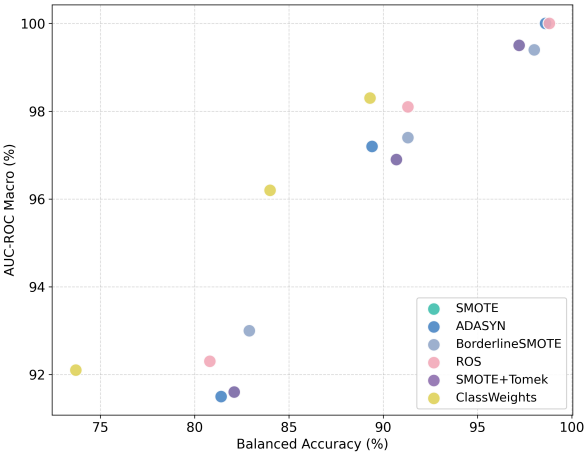


Fig. 9 Trade-off between macro AUC-ROC and balanced accuracy for every imbalance–kernel configuration.

4.5 Modified CNN with Batch Normalization

Augmenting the baseline architecture with batch normalization after each convolutional block(Figure 5) raises the overall accuracy to 99.41% with a cross-entropy loss of 0.228 after 60 epochs. The comparatively higher plateau loss relative to the baseline CNN reflects the regularizing effect of batch normalization, which favours well-generalised but less over-confident predictions; early validation-loss volatility (Figure 11) is consistent with the mismatch between per-batch and running BN statistics before convergence, stabilising after approximately epoch 20. The convergence is faster than for the baseline in modified CNN (Figure 13), with the training accuracy

Table 2 Top-ranked SVM configurations across the 6×3 ablation grid. Values are percentages.

Technique	Kernel	Bal. Acc.	Macro F1	Benign Prec.	Benign AUC
RandomOverSampler	RBF	98.81	98.81	98.82	100.00
ADASYN		98.61	98.63	98.31	100.00
Borderline-SMOTE		98.02	98.01	97.11	99.49
SMOTE		97.23	97.22	94.92	99.77
SMOTE–Tomek		97.23	97.22	94.92	99.77
Class weights		89.25	93.18	67.44	80.56
RandomOverSampler	poly	91.28	91.33	85.11	95.24
Borderline-SMOTE		91.28	91.35	85.48	94.64
SMOTE		90.69	90.71	83.77	95.24
SMOTE–Tomek		90.69	90.71	83.77	95.24

Table 3 Per-class performance of the best SVM model (RandomOverSampler + RBF kernel) on the test set.

Class	Precision	F1-score	Recall
Normal	0.98	0.98	0.99
Benign	0.99	0.99	1.00
Malignant	1.00	0.99	0.98

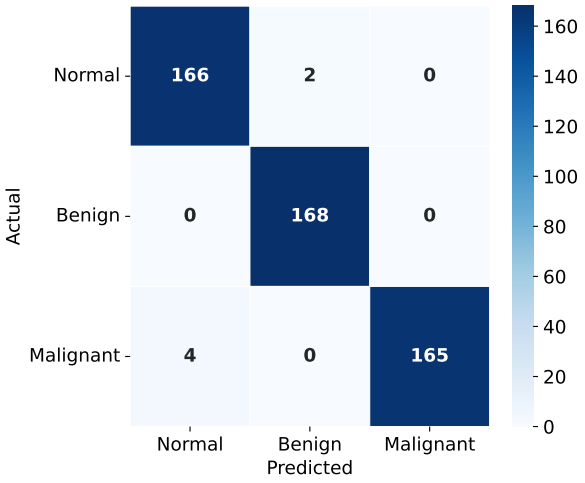


Fig. 10 Confusion matrix of the best SVM configuration (RandomOverSampler with RBF kernel).

rapidly approaching saturation and the validation accuracy stabilising near 99.4% within the first 20 epochs. Per-class metrics (Table 5) show perfect precision for benign and malignant cases and near-perfect performance on normal cases. The confusion matrix (Figure 14) confirms that of the 22 benign cases only one is mislabelled as

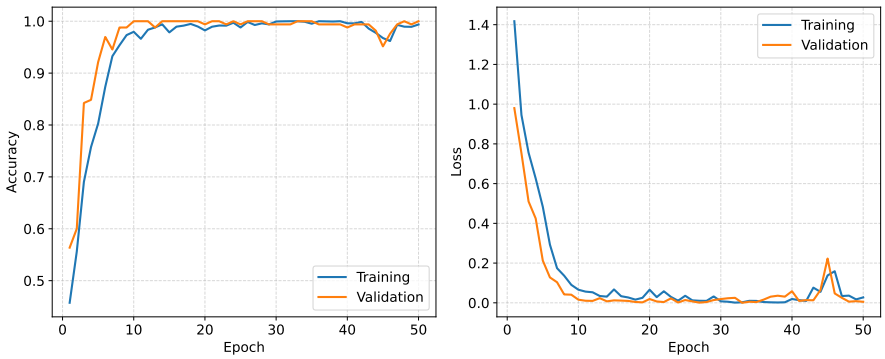


Fig. 11 Accuracy and loss curves for the baseline CNN over 50 training epochs.

Table 4 Per-class performance of the baseline CNN on the test set.

Class	Precision	F1-score	Recall
Benign	0.94	0.94	0.94
Malignant	1.00	0.99	0.99
Normal	0.98	0.98	0.98

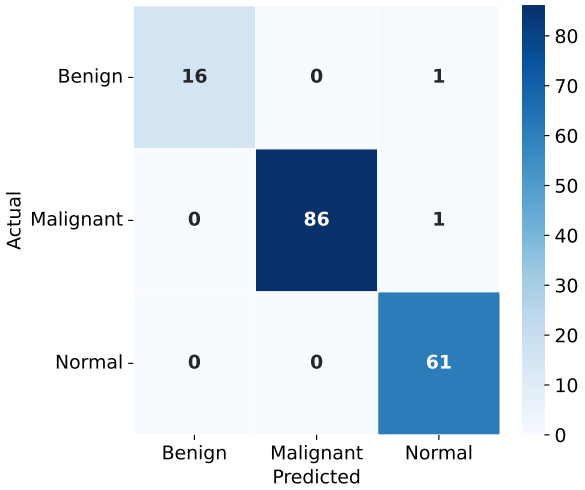


Fig. 12 Confusion matrix of the baseline CNN on the test set.

normal, all 85 malignant cases are correctly classified, and all 63 normal cases are correctly identified.

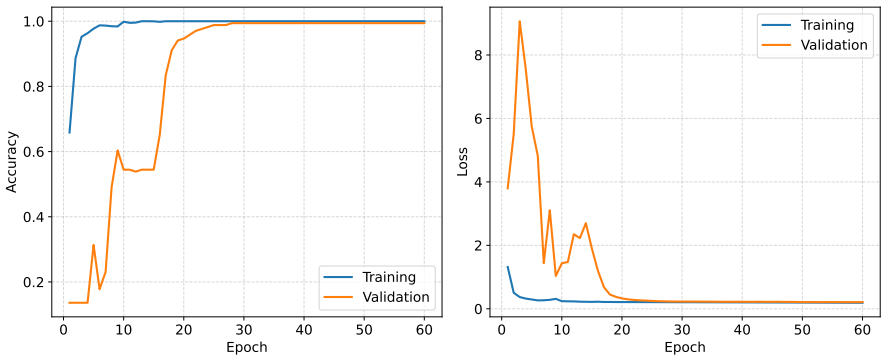


Fig. 13 Training and validation curves for the Modified CNN with batch normalization.

Table 5 Per-class performance of the modified CNN on the test set.

Class	Precision	F1-score	Recall
Benign	1.00	0.98	0.95
Malignant	1.00	1.00	1.00
Normal	0.98	0.99	1.00

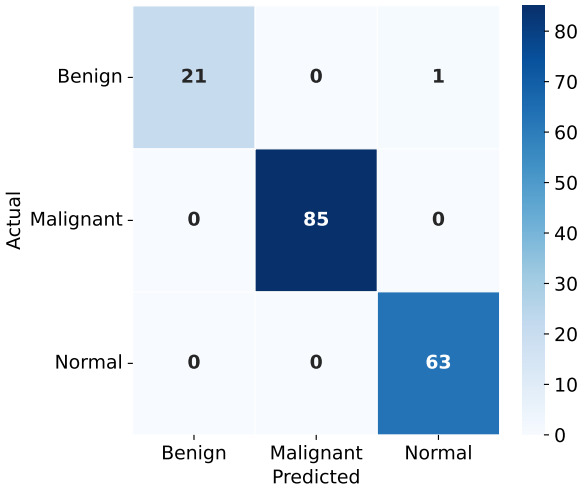


Fig. 14 Confusion matrix of the Modified CNN on the test set.

4.6 Hybrid CNN–SVM Model

The hybrid model employs EfficientNet-B0 (input size $224 \times 224 \times 3$) pre-trained on ImageNet with global average pooling, yielding a 1,280-dimensional embedding for each slice. A kernel SVM is then trained on the embeddings, with class balance enforced

Table 6 Per-class case counts before augmentation, after augmentation, and after SMOTE balancing within the hybrid pipeline.

Class	Original	After augmentation	After SMOTE(Train split only)
Normal	416	1 248	1 178
Benign	120	340	1 178
Malignant	561	1 683	1 178

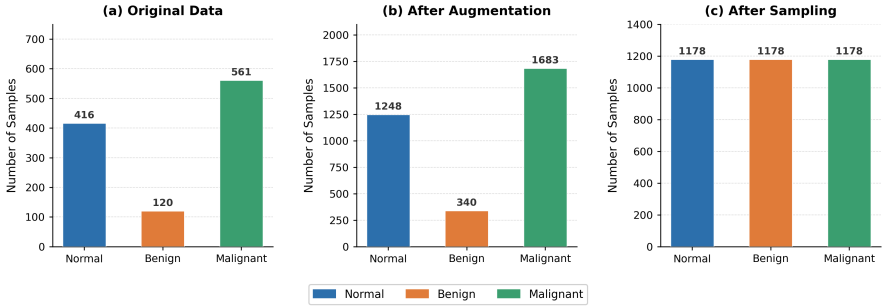


Fig. 15 Class distribution before augmentation, after augmentation, and after applying SMOTE within the hybrid CNN–SVM pipeline.

Table 7 Per-class performance of the hybrid CNN–SVM model on the validation set.

Class	Precision	F1-score	Recall
Normal	0.95	0.97	1.00
Benign	0.98	0.87	0.78
Malignant	0.99	0.99	1.00

through SMOTE and class weighting. Data augmentation expands the per-class counts and was applied to the full dataset prior to the train/validation/test split; SMOTE was subsequently applied only within the resulting training partition, equalizing all three classes to the size of the majority class in that partition (1,178) as summarised in Table 6 and visualised in Figure 15. After 25 fine-tuning epochs with early stopping, the model attains an overall test accuracy of 98.00% and a macro F1-score of 95.7%.

Validation-set metrics show perfect recall and precision on malignant cases, while benign cases attain a slightly lower F1-score of 0.87 due to the residual minority-class difficulty (Table 7, Figure 16). The hyperparameter tuning outcome (Table 8) selects an RBF kernel with $C = 10$ and $\gamma =$ auto, yielding a cross-validation accuracy of 99.66%. On the held-out test set (Table 9, Figure 17), benign-class recall rises while malignant-class metrics remain at 1.00.

The training and validation curves (Figure 18) corroborate the absence of overfitting: validation accuracy tracks training accuracy throughout, and the corresponding losses converge with only a marginal gap.

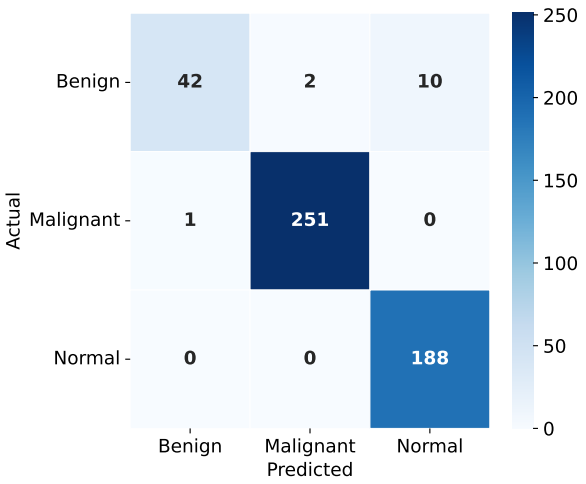


Fig. 16 Confusion matrix of the hybrid CNN–SVM model on the validation set.

Table 8 Hyperparameter-tuning results for the hybrid CNN–SVM classifier.

Parameter	Value
Class weights	$\{0 : 1.0,\ 1 : 1.0,\ 2 : 1.0\}$
Best parameters	$C = 10,\ \gamma = \text{auto},\ \text{kernel} = \text{RBF}$
Best cross-validation score	0.9966
Cross-validation accuracy	0.9966 ± 0.0042

Table 9 Per-class performance of the hybrid CNN–SVM model on the test set.

Class	Precision	F1-score	Recall
Normal	0.95	0.97	1.00
Benign	1.00	0.90	0.81
Malignant	1.00	1.00	1.00

4.7 Comparison of the Four Proposed Models

Table 10 consolidates the four proposed models for direct comparison. The Modified CNN provides the highest overall accuracy, while the SVM with RandomOverSampler and the hybrid CNN–SVM remain highly competitive and provide complementary advantages in terms of interpretability and computational cost.

4.8 Comparison with the State of the Art

Table 11 positions the Modified CNN against twelve representative published methods on lung-cancer detection from CT imagery, including very recent works. The proposed

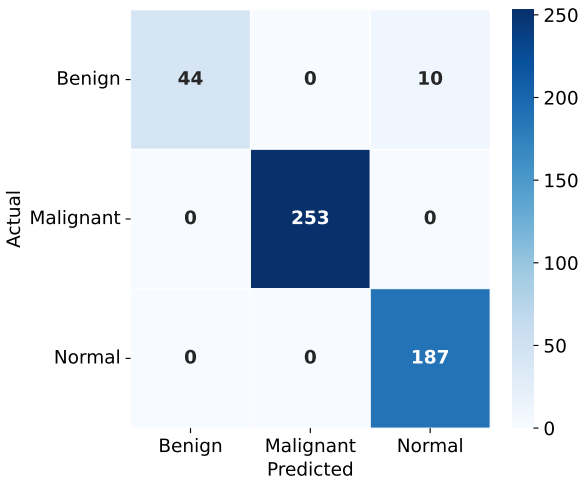


Fig. 17 Confusion matrix of the hybrid CNN–SVM model on the test set.

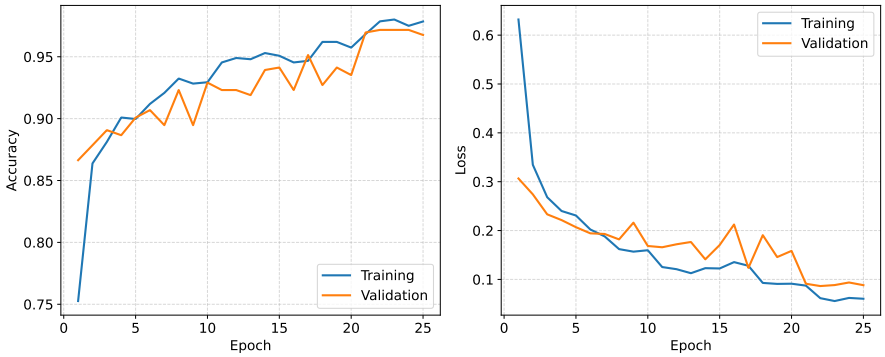


Fig. 18 Training and validation accuracy/loss curves for the hybrid CNN–SVM model.

Table 10 Summary of the four models on the IQ-OTH/NCCD dataset.

Model	Accuracy	Macro F1	Precision	Recall	MCC	Cohen's κ
SVM (ROS + RBF)	98.81	0.988	0.99	0.993	0.982	0.982
Baseline CNN	98.79	0.970	0.973	0.973	0.979	0.979
Modified CNN	**99.41**	**0.990**	**0.993**	**0.983**	**0.990**	**0.990**
Hybrid CNN–SVM (EffNet)	98.00	0.957	0.983	0.940	0.966	0.965

Modified CNN is competitive with the strongest recent benchmarks while requiring neither GAN-based augmentation nor transformer backbones. Compared with the

EfficientNet-B1-based Lung-EffNet [30] (99.10%), the Modified CNN attains an accuracy of 99.41% with a significantly simpler architecture, while the systematic 6×3 ablation grid contributes a methodological dimension.

5 Discussion

This paper proposed an integrated computer-aided framework for multi-class lung-cancer classification on the IQ-OTH/NCCD CT dataset. The framework consolidates a handcrafted preprocessing/segmentation pipeline, a four-category feature representation (statistical, LBP, gradient, shape), a systematic 6×3 ablation grid of class-imbalance strategies and SVM kernels, a baseline CNN, a Modified CNN with batch normalisation, and a hybrid EfficientNet–SVM model. The Modified CNN achieves the highest overall test accuracy of 99.41% with macro F1-score of 0.99; the SVM with RandomOverSampler and RBF kernel reaches 98.81% balanced accuracy; the hybrid model attains 98.00% test accuracy. Together, these results demonstrate that competitive performance can be obtained without resorting to large transformer backbones or GAN-based augmentation.

The dataset size remains the principal limitation, as IQ-OTH/NCCD encompasses only 110 patients drawn from a single institution. Deep-learning training is computationally expensive; the Modified CNN required several hours per training cycle on commodity GPU hardware, which may constrain real-time deployment without further hardware support. Multi-centre validation, standardisation of testing protocols, and adversarial robustness assessments remain to be carried out.

Five directions for future work are identified. First, dataset expansion through incorporation of LIDC-IDRI, LUNA16, and prospectively acquired multi-centre cohorts is expected to improve model generalisability across imaging protocols and patient populations. Second, hybrid spatio-temporal architectures combining CNN-based feature extraction with Bi-LSTM [12] or Transformer encoders may be investigated to capture inter-slice contextual information within volumetric scans. Third, visual explainability could be strengthened through Grad-CAM and SHAP overlays, quantitatively validated against radiologist-annotated regions of interest [13], thereby enhancing clinical interpretability and trust. Fourth, imbalance-aware loss formulations, including focal loss [22] and class-balanced softmax, represent a viable alternative to data-level resampling for mitigating class imbalance. Finally, uncertainty quantification through Monte-Carlo dropout or deep ensembles is proposed to yield calibrated confidence estimates, thereby supporting risk assessment in clinical decision-making.

Declarations

- **Funding.** No external funding was received for this research.
- **Conflict of interest / Competing interests.** The authors declare that they have no competing interests.
- **Ethics approval and consent to participate.** The IQ-OTH/NCCD dataset is publicly available and was originally collected with institutional review-board approval; no additional ethical approval was required for this secondary analysis.

Table 11 Summary of key studies in lung cancer diagnosis approaches with advantages and limitations.

Category	Reference	Method	Accuracy (%)	Remarks
SVM-based	Kareem et al. (2021) [15]	SVM + bit-plane slicing	89.88	Relies on manual bit-plane slicing; lowest accuracy among SVM baselines on this dataset
	Hassan et al. (2025) [28]	SVM vs. XGBoost/ANN/LR	99.30	No MCC or Cohen's κ reported – imbalance-aware performance unverified
	Hussain et al. (2022) [16]	SVM RBF/Poly + GLCM	–	Evaluated on MRI, not CT; no standard classification metrics reported
	This work	**SVM (ROS + RBF)**	**98.81**	**Evaluates six imbalance strategies across three kernels, with MCC, κ, and AUC-ROC reported**
Deep learning	Al-Yasriy et al. (2020) [11]	AlexNet CNN	93.54	Shallow backbone limits representational capacity
	Tasnim et al. (2024) [12]	Custom CNN + SMOTE	98.00	Only one imbalance strategy tested, with no comparative ablation
	Ruprah et al. (2024) [13]	VGG-16 transfer learning	99.00	Performance leans on ImageNet-pretrained weights
	Wadekar & Singh (2023) [29]	Modified VGG-19	97.73	Trained on histopathology images, not CT
	Shariff et al. (2025) [14]	CNN + differential aug.	98.78	Differential-augmentation pipeline adds training overhead
	Raza et al. (2023) [30]	Lung-EffNet (EffNet-B1)	99.10	Requires GAN-based augmentation
	Şenyiğit & Özbay (2025) [42]	LECNN (Goog/Eff/-Dark)	99.00	Three-backbone ensemble adds inference-time overhead
	This work	**Baseline CNN**	**98.79**	**Simple 3-block architecture with no augmentation, used as an internal reference point**
	This work	**Modified CNN**	**99.41**	**Comparable accuracy to recent benchmarks via one architectural change (batch norm), without GANs or ensembling**
Hybrid	Pawar et al. (2020) [27]	Decision tree + CNN	–	No accuracy reported, making direct comparison difficult
	Hasan et al. (2024) [32]	CNN + random forest	97.10	Lowest accuracy among hybrid approaches compared here
	Amouzad (2024) [33]	EfficientNetV2-S + SVM	97.89	Benchmarked on CIFAR-10 – not a lung-cancer or medical dataset
	Saikia et al. (2022) [17]	VGG + SVM/RF hybrid	98.70	Applies a single imbalance-handling technique
	This work	**Hybrid EffNet–SVM**	**98.00**	**Combines precomputed deep embeddings with an SVM decision function, supporting fast retraining**

- **Consent for publication.** Not applicable.
- **Data availability.** The IQ-OTH/NCCD dataset is publicly available on Mendeley Data and Kaggle.

  Mendeley link: https://data.mendeley.com/datasets/bhmdr45bh2/3

  Kaggle link: https://www.kaggle.com/datasets/hamdallak/the-iqothnccd-lung-cancer-dataset

- **Materials availability.** Not applicable.
- **Code availability.** Source code will be made available from the corresponding author upon reasonable request.
- **Author contribution.** S.A. and N.B.H. conceived and implemented the framework, conducted the experiments, and drafted the manuscript. R.I. supervised the work, provided methodological guidance, and revised the manuscript. All authors read and approved the final manuscript.

References

[1] American Cancer Society.: American Cancer Society Releases Latest Global Cancer Statistics; Cancer Cases Expected to Rise to 35 Million Worldwide by 2050. https://pressroom.cancer.org/GlobalCancerStatistics2024. American Cancer Society MediaRoom.

[2] The ASCO Post.: Cancer on the Global Stage: Incidence and Cancer-Related Mortality in Bangladesh. https://ascopost.com/issues/february-25-2017/cancer-on-the-global-stage-incidence-and-cancer-related-mortality-in-bangladesh/. ascopost.com.

[3] Ferlay J, et al.: Cancer Today. https://gco.iarc.who.int/today. International Agency for Research on Cancer, World Health Organization.

[4] American Cancer Society.: Lung Cancer Survival Rates: 5-Year Survival Rates for Lung Cancer. https://www.cancer.org/cancer/types/lung-cancer/detection-diagnosis-staging/survival-rates.html. cancer.org.

[5] Xie H, Yang D, Sun N, Chen Z, Zhang Y. Automated pulmonary nodule detection in CT images using deep convolutional neural networks. Pattern Recognition. 2019;85:109–119. https://doi.org/10.1016/j.patcog.2018.07.031.

[6] Giger ML, Chan HP, Boone J. Anniversary paper: history and status of CAD and quantitative image analysis: the role of medical physics and AAPM. Medical Physics. 2008;35(12):5799–5820. https://doi.org/10.1118/1.3013555.

[7] Biederer J, et al. Screening for lung cancer: does MRI have a role? European Journal of Radiology. 2017;86:353–360. https://doi.org/10.1016/j.ejrad.2016.09.016.

[8] Dandıl E. A computer-aided pipeline for automatic lung cancer classification on computed tomography scans. Journal of Healthcare Engineering. 2018;2018:1–12. https://doi.org/10.1155/2018/9409267.

[9] World Health Organization.: Lung Cancer. https://www.who.int/news-room/fact-sheets/detail/lung-cancer. WHO Fact Sheets.

[10] Huang X, Sun W, Tseng TL, Li C, Qian W. Fast and fully-automated detection and segmentation of pulmonary nodules in thoracic CT scans using deep convolutional neural networks. Computerized Medical Imaging and Graphics. 2019;74:25–36. https://doi.org/10.1016/j.compmedimag.2019.02.003.

[11] Al-Yasriy HF, Al-Husieny MS, Mohsen FY, Khalil EA, Hassan ZS. Diagnosis of lung cancer based on CT scans using CNN. IOP Conference Series: Materials Science and Engineering. 2020;928:022035. https://doi.org/10.1088/1757-899X/928/2/022035.

[12] Tasnim N, Noor KR, Islam M, Huda MN, Sarker IH. A deep learning based image processing technique for early lung cancer prediction. In: Proc. International Conference on Emerging Technologies for Sustainable and Intelligent Systems (ICETSIS); 2024. p. 1060–1064. https://doi.org/10.1109/ICETSIS61505.2024.10459494.

[13] Ruprah TS, Regmi B, Jadhav SB, Singh S. Early stage lung cancer detection using deep learning. In: Proc. MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon); 2024. p. 1–6. https://doi.org/10.1109/MITADTSoCiCon60330.2024.10575345.

[14] Shariff V, Paritala C, Ankala KM. Optimizing non small cell lung cancer detection with convolutional neural networks and differential augmentation. Scientific Reports. 2025;15(1):15640. https://doi.org/10.1038/s41598-025-98731-4.

[15] Kareem HF, Al-Huseiny MS, Mohsen FY, Khalil EA, Hassan ZS. Evaluation of SVM performance in the detection of lung cancer in marked CT scan dataset. Indonesian Journal of Electrical Engineering and Computer Science. 2021;21(3):1731–1738. https://doi.org/10.11591/ijeecs.v21.i3.pp1731-1738.

[16] Hussain L, et al. Lung cancer prediction using robust machine learning and image enhancement methods on extracted gray-level co-occurrence matrix features. Applied Sciences. 2022;12(13):6517. https://doi.org/10.3390/app12136517.

[17] Saikia T, Kumar R, Kumar D, Singh KK. An automatic lung nodule classification system based on hybrid transfer learning approach. SN Computer Science. 2022;3:272. https://doi.org/10.1007/s42979-022-01167-0.

[18] Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC

Genomics. 2020;21:6. https://doi.org/10.1186/s12864-019-6413-7.

[19] Gorodkin J. Comparing two K-category assignments by a K-category correlation coefficient. Computational Biology and Chemistry. 2004;28(5–6):367–374. https://doi.org/10.1016/j.compbiolchem.2004.09.006.

[20] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proc. 32nd International Conference on Machine Learning (ICML). vol. 37 of PMLR; 2015. p. 448–456. Available from: https://proceedings.mlr.press/v37/ioffe15.html.

[21] Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks. In: Proc. 36th International Conference on Machine Learning (ICML). vol. 97 of PMLR; 2019. p. 6105–6114. Available from: https://proceedings.mlr.press/v97/tan19a.html.

[22] Morgado J, et al. Machine learning and feature selection methods for EGFR mutation status prediction in lung cancer. Applied Sciences. 2021;11(7):3273. https://doi.org/10.3390/app11073273.

[23] Rehman A, Kashif M, Abunadi I, Ayesha N. Lung cancer detection and classification from chest CT scans using machine learning techniques. In: Proc. International Conference on Computer Applications & Information Security (ICCAIS); 2021. p. 101–104. https://doi.org/10.1109/CAIDA51941.2021.9425269.

[24] Madaan V, Sharma N, Chauhan R, Joshi K, Kumar BV, Sunil G. Classifying lung cancer disease using random forest algorithm. In: Proc. 4th International Conference on Innovative Practices in Technology and Management (ICIPTM); 2024. p. 1–7. https://doi.org/10.1109/ICIPTM59628.2024.10563203.

[25] Solyman S, Schwenker F. Lung tumor detection and recognition using deep convolutional neural networks. In: Communications in Computer and Information Science. Springer; 2023. p. 79–91. https://doi.org/10.1007/978-3-031-31327-1_5.

[26] Farhangi MM, Petrick N, Sahiner B, Frigui H, Amini AA, Pezeshk A. Recurrent attention network for false positive reduction in the detection of pulmonary nodules in thoracic CT scans. Medical Physics. 2020;47(5):2150–2160. https://doi.org/10.1002/mp.14076.

[27] Pawar VJ, Kharat KD, Pardeshi SR, Pathak PD. Lung cancer detection system using image processing and machine learning techniques. International Journal of Advanced Trends in Computer Science and Engineering. 2020;9(4):5956–5963. https://doi.org/10.30534/ijatcse/2020/260942020.

[28] Hassan M, Hassan S, Mujahid A, Umair M, Zubair M. Lung cancer diagnosis and classification using hybrid deep feature extraction. In: Proc. International Conference on Communication, Computing and Information Technology (ICCIT); 2025.

p. 256–261. https://doi.org/10.1109/ICCIT63348.2025.10989327.

[29] Wadekar S, Singh DK. A modified convolutional neural network framework for categorizing lung cell histopathological image based on residual network. Healthcare Analytics. 2023;4:100224. https://doi.org/10.1016/j.health.2023.100224.

[30] Raza R, Zulfiqar F, Khan MO, Arif M, Alvi A, Iftikhar MA, et al. Lung-EffNet: lung cancer classification using EfficientNet from CT-scan images. Engineering Applications of Artificial Intelligence. 2023;126:106902. https://doi.org/10.1016/j.engappai.2023.106902.

[31] Jozi NS, Al-Suhail GA. LCxNet: an explainable CNN framework for lung cancer detection in CT images using multi-optimizer and visual interpretability. Applied System Innovation. 2025;8(5):153. https://doi.org/10.3390/asi8050153.

[32] Hasan N, et al. Lung cancer classification using CNN and random forest. In: Proc. IEEE 3rd International Conference on Robotics, Automation, Artificial-Intelligence and Internet-of-Things (RAAICON); 2024. p. 276–281. https://doi.org/10.1109/RAAICON64172.2024.10928675.

[33] Amouzad A.: Classic machine learning: part 3/4 — classification on image dataset. https://medium.com/@a.r.amouzad.m/classic-machine-learning-part-3-4-classification-on-image-dataset-944ed3353d05. Medium.

[34] Otsu N. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics. 1979;9(1):62–66. https://doi.org/10.1109/TSMC.1979.4310076.

[35] Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002;24(7):971–987. https://doi.org/10.1109/TPAMI.2002.1017623.

[36] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research. 2002;16:321–357. https://doi.org/10.1613/jair.953.

[37] He H, Bai Y, Garcia EA, Li S. ADASYN: adaptive synthetic sampling approach for imbalanced learning. In: Proc. IEEE International Joint Conference on Neural Networks (IJCNN); 2008. p. 1322–1328. https://doi.org/10.1109/IJCNN.2008.4633969.

[38] Han H, Wang WY, Mao BH. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In: Lecture Notes in Computer Science. vol. 3644. Springer; 2005. p. 878–887. https://doi.org/10.1007/11538059_91.

[39] Batista GEAPA, Prati RC, Monard MC. A study of the behaviour of several methods for balancing machine learning training data. SIGKDD Explorations. 2004;6(1):20–29. https://doi.org/10.1145/1007730.1007735.

[40] Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995;20(3):273–297. https://doi.org/10.1007/BF00994018.

[41] Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement. 1960;20(1):37–46. https://doi.org/10.1177/001316446002000104.

[42] Şenyiğit B, Özbay E. A hybrid LECNN architecture: a computer-assisted early diagnosis system for lung cancer using CT images. International Journal of Computational Intelligence Systems. 2025;18:41. https://doi.org/10.1007/s44196-025-00761-3.