MV3Lung-NS: A Neutrosophic-Deep Learning Hybrid Framework for Computer-Aided Diagnosis of Chest X-Ray Scans

Abdelrhman Fathy Abdelrhman [a1*] ID, Ahmed R. Abas [a2] ID

[a] Department of Computer Science, Faculty of Computer and Informatics, Zagazig University, Zagazig 44519, Egypt
[1]asawmaa@fci.zu.edu.eg, [2]arabas@zu.edu.eg

Corresponding author E-mail: Abdelrhman Fathy Abdelrhman (*asawmaa@fci.zu.edu.eg*)

Abstract

The growing adoption of deep learning (DL) for chest X-ray (CXR) diagnosis faces three significant barriers that this study addresses. First, inherent ambiguities in CXRs - particularly overlapping tissue intensities between pathological and healthy regions, along with common noise artifacts - create indeterminate zones where conventional DL models frequently err. Second, the scarcity of high-quality annotated datasets and persistent class imbalance problems lead to biased and overfitted models. Third, the opaque decision-making process of DL systems undermines clinical trust, especially in borderline cases. To resolve these challenges, we implement Neutrosophic Sets (NS) to explicitly quantify and manage uncertainty at the pixel level through truth-falsity-indeterminacy memberships, particularly effective in clarifying ambiguous infection boundaries. Simultaneously, we employ radiologist-validated Data Augmentation (DA) techniques to mitigate data scarcity and imbalance issues. Our results demonstrate NS filtering enhances model reliability, improving EfficientNetB0 accuracy by 3.13% (94.79% to 97.92%) in uncertain regions, while DA boosts MobileNetV3Large's generalization capability with a 5.78% accuracy gain (93.75% to 99.53%). Building on these findings, we propose MV3Lung-NS, an integrated framework combining NS preprocessing, DA, and MobileNetV3Large that achieves state-of-the-art performance (99.53% accuracy, 99.65% precision) on pulmonary infection diagnosis. To bridge the interpretability gap, we implement Explainable AI (XAI) methods including SHapley Additive exPlanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME), and Gradient-weighted Class Activation Mapping (Grad-CAM), providing visual evidence that model decisions align with radiological markers of infection. This work makes dual contributions: advancing neutrosophic theory through empirical validation in medical imaging and delivering a clinically viable solution that addresses both technical and trust-related barriers in AI-assisted diagnosis.

Keywords: Lung-Infection Diagnosis; Deep Learning; Neutrosophic Sets; Data Augmentation; X-ray Scans, Explainable Artificial Intelligence.

1. Introduction

Lung infections are a major cause of global illness and death [1]. They can result from various pathogens, leading to different types of pulmonary infections. Bacterial infections, often responsible for pneumonia, are commonly linked to pathogens like Streptococcus pneumoniae, which require antibiotics and imaging for diagnosis [2]. Viral infections, such as those caused by influenza, can also lead to pneumonia, with specific imaging techniques aiding in diagnosis [2]. Mycobacterial infections from Mycobacterium tuberculosis result in pulmonary tuberculosis and require specialized imaging for proper diagnosis [2]. Fungal infections, particularly those caused by Aspergillus, usually affect

immunocompromised individuals [2]. Aspiration pneumonia occurs when food or liquid is inhaled, while community-acquired pneumonia (CAP) is contracted outside healthcare settings, often due to bacteria or viruses [2]. This emphasizes the necessity for prompt diagnosis and treatment [2]. Pneumonia significantly contributes to global morbidity and mortality, impacting all age groups, especially in low- and middle-income countries. It is a leading cause of death, and the Coronavirus Disease 2019 (COVID-19) pandemic has further underscored its serious impact [2]. COVID-19 was caused by the new severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) and was first diagnosed in China on December 31, 2019[3]. SARS-CoV-2 spread quickly over the world, and on January 30, 2020, the World Health Organization (WHO) proclaimed the COVID-19 outbreak a public health emergency of worldwide concern [4]. The number of infected cases and fatalities from COVID-19 is rapidly increasing, with over 14,765,256 confirmed cases and over 612,054 deaths across 200 countries as of July 22, 2020, [5]. The crisis presents unprecedented challenges and requires significant efforts from governments, healthcare professionals, and individuals to mitigate its effects [6].

Medical imaging is essential for diagnosing diseases and planning treatments by allowing non-invasive visualization of internal structures [7]. Technologies such as chest X-rays (CXR), computed tomography (CT), magnetic resonance imaging (MRI), Positron Emission Tomography (PET) ,and ultrasound have transformed healthcare by providing valuable insights into anatomical and functional abnormalities [8]. These imaging methods facilitate early disease detection, monitor disease progression, assess treatment effectiveness, and guide surgical and interventional procedures [7]. Among the various imaging techniques available, CXR and CT scans are the most commonly used to examine patients with Lung infections[9, 10]. CT scans and CXR differ in technology, detail, and clinical applications. CT scans utilize X-rays and computer processing to create high-resolution 3D cross-sectional images, which provide superior contrast and enable better detection of lung abnormalities, such as ground-glass opacities. In contrast, chest X-rays produce 2D images with less detail, making them less sensitive to early or subtle lung changes. Moreover, CT scans involve higher radiation exposure and require specialized equipment, while X-rays are faster, more accessible, and use lower radiation doses. For these reasons, chest X-rays are often the preferred choice for initial screenings and routine checks, particularly in resource-limited settings. CT scans are typically reserved for more severe or complex cases that necessitate detailed assessment [11]. Medical images can be affected by various types of noise, such as Gaussian and Poisson noise, which are frequently observed in CXR and CT scans [12]. Poisson noise, in particular, can make it difficult for clinicians to identify bone fractures in X-ray images [13, 14].

Recently, Deep Learning (DL) technologies have demonstrated significant potential in processing medical images [15]. DL models are able to process a large number of features that represent unstructured data [16]. However, two critical unresolved challenges impede their use in clinical adoption:

1. **Aleatoric Uncertainty**: Noise in medical image acquisition (e.g., sensor noise, motion blur) introduces irreducible uncertainty [12]. Traditional denoising techniques, such as wavelet filters, Gaussian smoothing, and anisotropic diffusion, often over-smooth features or amplify artifacts, complicating disease identification [17]. For deblurring, methods like the Richardson-Lucy algorithm, Wiener filter, and regularized filter can effectively recover images but may also introduce noise amplification and boundary artifacts [18] Additionally, contrast enhancement techniques, including normalization and histogram equalization, may result in unnatural brightness and further amplify noise [17]

2. **Epistemic Uncertainty**: refers to the uncertainty related to model parameters. It arises from the limited ability of observed values to fully capture the true underlying phenomena, highlighting the difference between the predicted value $\hat{y}$ and the actual value y. This type of uncertainty can be quantified using mutual information[19].

To address these challenges, this paper introduces a framework, referred to as MV3Lung-NS (MobileNetV3Large, Lung, Neutrosophic Set), for lung-infection diagnosis using CXR scans. a novel framework that integrates:

- Pre-trained Model
  MobileNetV3Large [20] was chosen after comparing it to six other pretrained models: MobileNetV3Small [20], ResNet50V2[21], EfficientNetB0[22], InceptionV3[23], NASNetMobile[24], and MobileNetV2[25].

- Neutrosophic Set

  NS to model indeterminacy in CXR labels and features, reducing aleatoric uncertainty. This integration of NS into deep learning models significantly improves their performance, as evidenced by the comparison of metrics' values before and after NS utilization. For instance, the EfficientNetB0 model achieves a substantial increase in accuracy (from 0.9479 to 0.9792), precision (from 0.9562 to 0.9785), recall (from 0.9062 to 0.9583), and F1 score (from 0.9303 to 0.9683). Other models, such as the ResNet50V2 and the MobileNetV3Small, also show improvements, demonstrating the robustness and reliability of NS in handling uncertainty and imprecision in data.

- Data Augmentation

  Data augmentation (DA) enhances a model's generalization and resilience by exposing it to various data variations, reducing overfitting [53]. Ensuring balanced classes in the dataset promotes fair representation during training, preventing bias toward majority classes and facilitating effective learning from minority samples.

While DA and selected a specified pretrained model (MobileNetV3Large) mitigate epistemic uncertainty by improving data coverage and feature reuse, future work could integrate Bayesian layers to explicitly quantify model confidence

- Stochastic Gradient Descent with Warm Restarts

  Stochastic Gradient Descent with Warm Restarts (SGDR) algorithm is used for optimizing training and utilizing XAI for interpreting results. The MV3Lung-NS framework achieves superior performance by leveraging the strengths of NS and DA while ensuring efficient model training and convergence together with interpretation of results. The proposed framework overcomes the limitations of other DL models shown in the literature and achieves the following contributions:

- Interpreting results through utilizing XAI

  The framework incorporates XAI techniques such as SHapley Additive exPlanations (SHAP)[26], Local Interpretable Model-agnostic Explanations (LIME)[27], and Gradient-weighted Class Activation Mapping (Grad-CAM) [28], to enhance transparency and interpretability of results. SHAP provides feature importance scores. LIME offers local interpretability. Grad-CAM generates heatmaps to visualize decision-making regions in the images. These methods ensure that the framework's predictions are understandable and trustworthy, particularly in critical applications like healthcare.

The NS-enhanced diagnosis significantly improves accuracy, as evidenced by a 3.1% increase in performance for EfficientNetB0 due to the resolution of label ambiguity. Additionally, data augmentation (DA) has enabled MobileNetV3Large to achieve an impressive accuracy of 99.5%, compared to a baseline of 93.8%. This approach also introduces explainable predictions by combining NS uncertainty maps with Grad-CAM heatmaps, ensuring that the interpretations align with the reasoning of radiologists. The MV3Lung-NS framework achieves state-of-the-art performance, boasting an accuracy of 99.53% and a precision of 99.65% in diagnosing pulmonary infections.

The following sections of this paper are organized as follows: Section 2 discusses the concepts of Neutrosophic Sets and provides their definitions. Section 3 outlines the related work that utilizes Neutrosophic Sets and Deep Learning algorithms for image classification. Section 4 describes the proposed framework. Section 5 presents the experimental results. Section 6 presents discussion of results. Finally, Section 7 concludes the paper and introduces suggestions for future work.

2. Neutrosophic Sets

Florentin Smarandache introduced the Neutrophilic Set (NS), a broad dominant framework that generalizes the fuzzy set, classical set, intuitionistic fuzzy set, interval-valued fuzzy set, and interval-valued intuitionistic fuzzy set models [29]. Some examples of neutrosophic techniques include neutrosophic probability, neutrosophic set theory, neutrosophic logic, and neutrosophic statistics. NS addresses issues of indeterminacy and uncertainty across various fields. In neutrosophic theory, each event is characterized by distinct degrees of truth, falsehood, and indeterminacy, which can be evaluated independently of one another[29]. The indeterminate value addresses uncertainties. The neutrosophic image properties allow the NS to achieve superior performance in several image denoising applications in computer vision and image processing [29].

Neutrosophy extends both fuzzy and classical set theories. The neutrosophic theory is related to neutrosophy, which is applied across various specializations to address problems involving indeterminacy and uncertainty. Neutrosophy introduces new mathematical theories that enhance both classical and fuzzy approaches. Examples of neutrosophic techniques include neutrosophic statistics, neutrosophic logic, neutrosophic set theory, and neutrosophic probability. In neutrosophy theory, each event is assigned a specific degree of truth, falsehood, or indeterminacy that can be considered independently.

Neutrosophic sets provide a foundation for several types of sets, including classic sets, intuitionistic fuzzy sets, interval-valued fuzzy sets, fuzzy sets, paraconsistent sets, paradoxist sets, interval-valued intuitionistic fuzzy sets, dialetheist sets, and tautological sets [30–32]. Neutrosophic sets are defined in various ways [33, 34]:

2.1 Neutrosophic Concepts

2.1.1 Neutrosophic sets (NS)

In neutrosophic sets, each element is characterized by three membership functions: truth-membership (T), indeterminacy-membership (I), and falsity-membership (F), which are independent of each other and can take values within the interval [0, 1]. Mathematically, neutrosophic sets **S** in a universe **W** are represented in Equation 1 [35].

$$S = \left\{ \left[w, \left(T_s(w), I_s(w), F_s(w) \right) \right], w \in W \right\} \qquad (1)$$

Where $T_s(w)$, $I_s(w)$, $F_s(w) \in [0,1]$ and $0 \leq T_s(w)+I_s(w)+F_s(w) \leq 3$.

2.1.2 Neutrosophic images

In neutrosophy, an image from the spatial domain is conceptualized as a neutrosophic image (NI). Let Z represent the universe of discourse and W = w* w denotes the image window, which includes the rows and columns of the spatial domain. Thus, W is a collection of image intensity pixels, where W is a subset of Z that comprises bright pixels [36].

A neutrosophic image is generally defined by three membership sets: T, I, and F, as indicated in Equation 2 [35]. For an image with dimensions A * B, each pixel P (a, b) in the neutrosophic image is

represented as P$_{NS}$ (a, b). The values P$_{NS}$ (a, b) reflect the memberships of bright, indeterminate, and dark intensity values through the true membership T (a, b), indeterminate membership I (a, b), and false membership F (a, b). These relationships are further defined by Equations 3-5 [35].

$$P_{NS}(m,n) = \{T(a,b), I(a,b), F(a,b)\} \qquad (2)$$

$$T(a,b) = \frac{\overline{g}(a,b) - \overline{g}_{min}}{\overline{g}_{max} - \overline{g}_{min}} \qquad (3)$$

$$F(a,b) = 1 - T(a,b) \qquad (4)$$

$$I(a,b) = \frac{\delta(a,b) - \delta_{min}}{\delta_{max} - \delta_{min}} \qquad (5)$$

where $\overline{g}$ (a, b) represents the local mean value of the image, and δ (a, b) is the absolute value defined as the difference between the intensity and the $\overline{g}$ (a, b) and calculated thought Equations 6 and 7 [35]:

$$\overline{g}(a,b) = \frac{1}{w*w} \sum_{m=i-w/2}^{i+w/2} \sum_{n=j-w/2}^{j+w/2} g(m,n) \quad (6)$$

$$\delta(a,b) = abs\big(g(a,b) - \overline{g}(a,b)\big) \qquad (7)$$

2.1.3   Neutrosophic image entropy

The entropy of a grayscale image reflects the distribution of its intensity levels. When entropy is at its maximum, it indicates that the intensity values are equally likely and spread out evenly. In contrast, low entropy suggests that the probabilities of different intensity levels vary, leading to a non-uniform distribution [37].

Neutrosophic image entropy is defined as the sum of the entropies of three sets: T, I, and F. This metric evaluates the distribution of elements within the neutrosophic domain and calculated through the following Equations 8-11 [29].

$$En_{NS} = En_T + En_I + En_F \qquad (8)$$

$$En_T = - \sum_{i=\min\{T\}}^{\max\{T\}} H_T(i) \ln H_T(i) \qquad (9)$$

$$En_I = - \sum_{i=\min\{I\}}^{\max\{I\}} H_I(i) \ln H_I(i) \qquad (10)$$

$$En_F = - \sum_{i=\min\{F\}}^{\max\{F\}} H_F(i) \ln H_F(i) \qquad (11)$$

In this context, En_T, En_I, and En_F represent the entropies of the sets T, I, and F, respectively. Meanwhile, $H_T(i)$, $H_I(i)$,and $H_F(i)$ denote the probabilities associated with the elements in the sets T, I, and F, respectively.

2.2 Neutrosophic Framework for Image Processing

Neutrosophic sets provide considerable benefits in the analysis of medical images by adeptly handling inconsistencies, indeterminacies, and uncertainties [38]. This section outlines the comprehensive framework encompassing all crucial stages of Neutrosophic Logic, Image preprocessing, feature extraction, and classification. This ability enhances reasoning and computational processes, making neutrosophic expert systems particularly effective for research in medical imaging. Techniques based on neutrosophic sets support various image-processing tasks, including thresholding, denoising, clustering, segmentation, and classification, across different medical imaging modalities. These methods improve the accuracy and reliability of interpreting medical images and effectively address the complexities of medical data [39].

2.2.1   Noise Reduction and Enhancement

Preprocessing is a vital initial step in neutrosophic image-processing techniques. This stage enhances the quality of medical images by reducing noise and highlighting key features. In the neutrosophic domain, noise reduction is attainable through the appropriate application of filters, which can effectively modify the image's structure and ultimately lead to reduced noise levels [40].

For a specific image S (x, y), the transformation can be expressed as outlined in Equation 2. In this context, noise reduction involves adjusting these elements to minimize uncertainty as much as possible, while still preserving the integrity of true and false values. This can be articulated by using the following equations:

$$T(a,b) = T'(a,b) \qquad\qquad (12)$$

$$I(a,b) = I'(a,b) * G(a,b) \qquad\qquad (13)$$

$$F(a,b) = F(a,b) \qquad\qquad (14)$$

where G (a, b) is a Gaussian filter applied to the indeterminacy component to reduce noise.

Enhancement is modification of truth component to increase relative contrast and display of critical features. This can be achieved using techniques like histogram equalization or contrast stretching on T (a, b) [41].

2.2.2   Feature Extraction

Feature extraction involves converting the raw pixel values from an image into a collection of features, which are typically distinctive characteristics of input patterns. These features can be utilized in selection and classification tasks. Techniques for feature extraction are generally categorized into four types: geometrical, statistical, model-based, and signal processing [42].

Neutrosophic logic improves feature extraction by considering varying degrees of truth, indeterminacy, and falsehood. When employing neutrosophic logic techniques for feature extraction, various statistical measures—including mean, variance, and entropy—can be derived from these neutrosophic components. These measures not only highlight uncertainties and inconsistencies within the image but also offer insights that extend beyond mere intensity distribution. A notable application of this approach is in the detection and diagnosis of diseases, such as breast cancer using mammogram images. In this context, the neutrosophic features extracted to train a machine learning model achieved an accuracy of

95% in distinguishing malignant tumors from benign ones, surpassing the 88% accuracy attained using conventional feature extraction methods[39].

3   Related Work

This section reviews the relevant literature on the use of NS in conjunction with DL techniques for image classification. The significance of NS in DL models is examined by utilizing a limited dataset of COVID-19 X-ray images [43]. The images are transformed into the NS domain, which comprises three categories: true (T) images, indeterminate (I) images, and false (F) images. The obtained images are then employed to train a test various DL models including: AlexNet [44], GoogleNet[45], and ResNet18 [46]. Performance of these models is assessed in terms of accuracy, precision, recall, and F1 score. Results indicate that integrating NS with DL models enhances testing accuracy, particularly given the limited availability of COVID-19 datasets. The Indeterminacy (I) NS domain achieves the highest testing accuracy that is 87.1%.

Cai et al.[47] introduce an automated method for detecting clustered microcalcifications (MCCs) that employs NS domain transformation, similar to a traditional computer-aided detection (CAD) system. A deep convolutional neural network (DCNN) classifier is developed to identify individual microcalcifications while minimizing false positives. Additionally, a neutrosophic reinforcement sample learning (NRSL) technique is implemented to accelerate the learning process. An evaluation study based on clustering shows that the MCC detection method has a sensitivity of 92.5% at a rate of 0.50 frames per second per image. A robust deep convolutional neural network classifier is also created for diagnosis using automatic detection, achieving area under the curve (AUC) values of 0.908 and 0.872, respectively. Performance results indicate that the proposed method significantly enhances the automated detection and classification of MCCs in full-field digital mammograms.

Guo and Ashour [48] propose a classification framework consisting of two phases: multiple deep convolutional neural networks (MDCNN) and NS approach. The NMDCNN determines the number of reinforced training sessions for each epoch with the aid of NS and then classifies dermoscopic images as either malignant or benign using incremental learning and maximum voting. The effectiveness of this model is evaluated using the International Skin Imaging Collaboration dataset.

In the field of brain tumor segmentation, NS approaches are integrated with deep learning (DL) to identify tumor regions in MRI scans [49]. NS is employed to address uncertainties in tissue boundaries and combined with convolutional neural networks (CNNs) to boost segmentation precision. Results indicated that preprocessing MRI scans with NS significantly improved the performance of DL models, especially in situations involving unclear or noisy data.

Yasser et al. [50] developed an effective and user-friendly diagnostic method for the automatic detection of COVID-19 through digital chest X-rays. This tool utilizes a hybrid architecture that integrates nature-inspired approaches and machine learning (ML). It extracts classification characteristics from X-ray images by utilizing morphological features (MF) and principal component analysis (PCA). The machine learning networks differentiate between chest X-rays of patients positive for COVID-19 and those of healthy individuals.

NS are utilized for the detection of colorectal polyps in colonoscopy images [51]. A saliency detection network augmented with NS is put forward to enhance the identification of polyp areas. By

establishing T, I, and F functions, ambiguity is successfully minimized, and the precision of polyp detection is boosted, resulting in a high accuracy rate.

A novel methodology [35] is presented for differentiating between various types of lung infections, including COVID-19, bacterial pneumonia, and viral pneumonia, using CXR scans. The authors identify the challenges posed by the fuzziness and imprecision inherent in medical imaging data, which can complicate accurate diagnoses. To address these issues, the study employs a neutrosophic set approach that classifies image data into True (T), False (F), and Indeterminacy (I) memberships. This method enhances feature extraction and reduces uncertainty in the classification process. The methodology involves preprocessing the images through alpha-mean and beta-enhancement operations to improve the clarity of lung opacity levels before inputting them into DL models such as ResNet-50, VGG-16, and XGBoost. Experimental results demonstrate that the enhanced neutrosophic images achieve an accuracy of 97.33%, outperforming other domain sets.

Guo et al. [52] develop a deep neural network (DNN) designed to extract white blood cells (WBCs) from blood images, addressing the challenge of object indeterminacy in the nuclear segmentation domain. The network integrates WBC indeterminacy as a fusion element to enhance segmentation into distinct areas. This model surpasses three original encoder-decoder networks, achieving high precision rates and the highest mean segmentation accuracy (95.3%).

Abdullah [53] proposes a hybrid method combining NS theory with deep learning to enhance medical X-ray classification under uncertainty, like noise and ambiguity. The approach transforms images into three NS domains—True (T), Indeterminacy (I), and Falsity (F)—improving feature representation. Evaluated on datasets for cervical spine injuries and chest diseases, models like DenseNet121 and MobileNet showed better performance. DenseNet121 achieved 99.67% accuracy for spinal fractures and 88.40% for chest diseases, both in the T domain.

Elbehiery et al. [54] present an advanced machine learning framework for breast cancer detection that utilizes NS to manage uncertainty and vagueness in medical data. The study transforms the original breast cancer dataset into neutrosophic representations, incorporating truth, indeterminacy, and falsity components to enhance classification accuracy. Four machine learning models—logistic regression, support vector machine (SVM), gradient boosting, and k-nearest neighbors (KNN)—are evaluated with and without neutrosophic preprocessing. The results indicate that models enhanced with neutrosophic processing achieve superior performance, with logistic regression reaching the highest accuracy of 98.6%, compared to 95.8% for its non-neutrosophic version.

Table 1 describes the related works through the year, task, disease, modality, dataset, number of images, number of classes, model, and obtained accuracy. This table shows that DL applications of NS are still on the rise. Therefore, this paper presents a novel framework that integrates the NS and DL environments to classify Lung infections using CXR scans.

Table 1 Description of related works that utilize NS and DL for the analysis of medical images.

Reference	year	Task	Disease	Modality	Dataset	No. images	No classes	Model	Accuracy
[47]	2019	Cluster classification	Breast cancer	Mammograms	NFH dataset	676	2	DCNN	81.3%

[48]	2019	classification	Skin cancer	Dermosco-basepic images	c ISIC2016	1279	2	MDCNN	85.22%
[49]	2019	Segmentation, Classification	Brain tumor	MRI	TCIA	500	2	CNN	95.62%
[43]	2021	Classification	COVID-19	X-ray	COVID-19 x-ray dataset	306	4	AlexNet, GoogleNet, Restnet1	87.1% for I domain
[50]	2022	Classification	COVID-19	X-ray	COVID-19 Dataset, healthy dataset	570	2	(MFs), (PCA)	98.46%
[51]	2022	Segmentation	Colorectal polyp	-	EndoScene, Kvasir-SEG	EndoScene= 912 Kvasir-SEG=1000	EndoScene =8 Kvasir-SEG =4	saliency detection network	EndoScene =0.971
[35]	2023	Classification	COVID-19	X-ray	ActualMed COVID-19 Chest X-ray and COVID-19 radiography dataset	360	4	XG-Boost	97.33%
[52]	2024	Segmentation	-	pathological imaging	JTSC, CellaVision, SegPC	300, 100, 2633	5	Encoder-Decoder	95.3%
[53]	2025	Classification	Fracture, Dislocation, Normal	X-ray	Cervical Spine X-ray Dataset	2,009	3	DenseNet121 (T) InceptionV3 (T)	99.67%
[53]	2025	Classification	Hemothorax, Pneumothorax, Flail Chest, Normal	X-ray	Chest Disease X-ray Dataset	1,950	4	DenseNet121 (T)	88.40%
[54]	2025	Classification	Breast Cancer	X-ray	Breast Cancer Dataset (Kaggle subset)	570	2	Logistic Regression (N-data)	98.6%

4 Proposed framework

In this section, we discuss the proposed framework, MV3Lung-NS, which is based on NS, DA, various DL models, and XAI.

4.1 Data preprocessing steps

(a) Resizing images

Resizing images is an important preprocessing step in DL models, especially for image analysis tasks such as classification and object detection. These models often require fixed input dimensions, making resizing essential for standardizing images and reducing computational load. Proper resizing is vital to ensure optimal model performance and efficiency [55]. In this research, all images are resized to a fixed size of 224 x 224 pixels.

(b) Converting RGB X-ray images into grayscale images.

In RGB images, each pixel is represented in a 24-bit color space. This pixel is converted to an 8-bit grayscale space with values ranging from 0 to 255 in grayscale images.

4.2 Applying Neutrosophic filtering

Neutrosophic filtering is used for handling uncertainties resulting from different sources of noise and distortions in grayscale CXR scans. Three subsets have resulted from this filtering step, see equations [3-5], that are the components of the Neutrosophic set: Truth (T), Indeterminacy (I), and Falsity (F). Each component is processed individually and fed into deep learning models to evaluate their impact on image quality and model performance.

4.3 Applying data augmentation methods

The DA methods applied in this research include brightness adjustment, and image flipping. These methods are used to enlarge the size of the input dataset to improve classification performance of the DL models. DA results in:

- Diversifying the input dataset, which increases the model's generalization capability and improves resilience. This exposure to various data variations helps reduce overfitting [56].
- Ensuring balanced classes in the input dataset, which is results in fair representation during training. This balance prevents bias towards the majority classes and allows for effective learning from minority class samples.

4.4 Classification model

After preprocessing CXR scans, they are used as input for various pretrained DL models, including: MobileNetV3Large, MobileNetV3Small, ResNet50V2, EfficientNetB0, InceptionV3, NASNetMobile, and MobileNetV2. These models are evaluated based on their performance in the diagnosis of lung infection in these CXR scans into three categories: COVID-19, Viral Pneumonia, and Normal Chest.

5 Experiments and results

5.1 The input dataset

The dataset used in the experiments is published on Kaggle website [57]. It includes 317 CXR scans categorized into three classes: COVID-19, Viral Pneumonia, and Normal. The dataset contains 137 scans of COVID-19, along with 90 scans of Viral Pneumonia and 90 scans of Normal CXR scans. Table 2 presents a description of the dataset. Samples of the training images from the dataset are shown in Figure 1.

Table 2. Description of the input dataset.

Infection Type	Number of scans
COVID-19	137
Normal	90
Viral Pneumonia	90

(A) COVID-19   (B) Normal   (C) Viral Pneumonia.
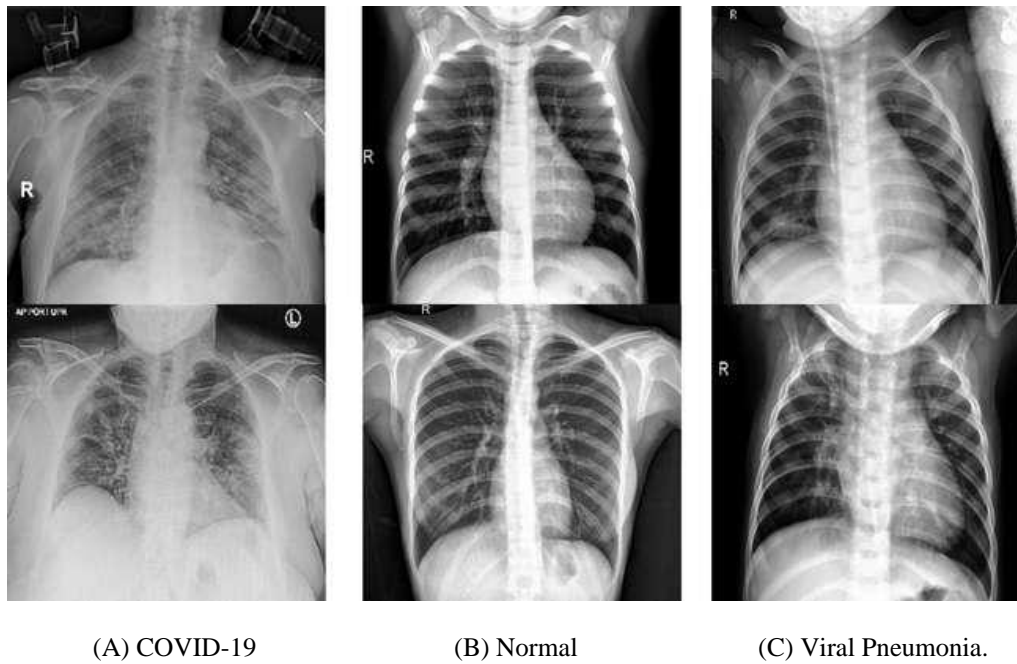
Figure 1. Samples of the scans from the input dataset with their true diagnosis (A) Covid-19, (B) Normal, and (C) Viral Pneumonia.

5.2 Hyperparameter Optimization using GridSearchCV

In this research, we employ the grid search cross-validation (GridSearchCV)[58] to improve the model's performance. We used the GridSearchCV with different optimizers including: Adam, SGD, and RMSprop, to test various combinations of hyperparameters in our experiments to identify the best set. Table 3 presents the hyperparameters used in the GridSearchCV. Hyperparameters are predefined values established before training the model, and their combinations are assessed in a grid format to evaluate the model's effectiveness. For example, when working with two hyperparameters, we evaluate the model's performance by changing one hyperparameter while keeping the other constant. This process is then reversed to ensure a thorough assessment. This systematic approach allows for a comprehensive exploration of the specified parameter space, ultimately leading to the discovery of the most effective parameter combination for the model.

Additionally, we incorporate the k-fold cross-validation [59] into the GridSearchCV methodology, setting k to 3. This means the dataset is split into Three subsets, with cross-validation performed on each one. To evaluate the performance of each hyperparameter combination, we use the Accuracy measure. Table 4 presents the optimal hyperparameters for each pre-trained model. This approach facilitates the selection of the optimal hyperparameter combination, which is subsequently applied to the entire dataset to create a new model.

Table 3. The Hyper-Parameters that are used for the GridSearchCV with different optimizers

Batch size	Learning rate
12	0.1
16	0.01
32	0.001

64	0.0001

Table 4. The optimal Hyper-Parameters for each pre-trained model

Hyper parameter	MobileNetV3Large	MobileNetV3Small	ResNet50V2	EfficientNetB0	InceptionV3	NASNetMobile	MobileNetV2
Optimizer	**Adam**	Adam	Sgd	Adam	RMSprop	sgd	sgd
Batch size	**12**	16	16	12	12	12	12
Accuracy	**93.48 %**	93.45%	85.65%	93.46%	87.92%	83.08%	90.27%
Learning rate	**0.01**	0.01	0.1	0.01	0.01	0.001	0.1

5.3 Stochastic Gradient Descent with Warm Restarts (SGDR)

Stochastic Gradient Descent with Warm Restarts (SGDR) algorithm is used for adjusting learning rate (LR) coefficient during different training epochs [60]. The SGDR algorithm gradually decreases the LR through the consequent training epochs, which ensures effective and efficient LR updates during training [61]. Table 5. describes the Key parameters of the SGDR algorithm.

Table 5. Key parameters of the SGDR algorithm.

SGDR parameter	Value
Min LR	1e-6
Max LR	1e-3
LR Decay	0.9
Cycle length	10
Mult Factor	2

5.4 Hyperparameter Determination Based on GridSearchCV and SGDR

The hyperparameters listed in Table 6 were carefully selected through a two-stage optimization process: (1) an exhaustive GridSearchCV (described in Section 5.2) to identify the best-performing static hyperparameters, followed by (2) SGDR (described in Section 5.3) to dynamically fine-tune the learning rate during training.

From the GridSearchCV results, the Adam optimizer and a batch size of 12 consistently achieved the highest accuracy, with MobileNetV3Large reaching 93.48%. This configuration also demonstrated stable convergence. Although some models, such as ResNet50V2, performed well with alternative settings (like SGD), the combination of Adam and a batch size of 12 was selected as the default optimizer due to its robustness across most tested architectures.

To further improve training efficiency, SGDR was utilized to dynamically adjust the learning rate within the bounds of 1e-6 (minimum learning rate) and 1e-3 (maximum learning rate). The decay factor

was set to 0.9, with a cycle length of 10 epochs. This approach ensured that the model could avoid local minima and maintain steady progress without requiring manual learning rate scheduling.

Table 6. The DL models training hyperparameters.

Hyper parameter	Value
Loss function	categorical cross-entropy
Optimizer	Adam
Metrics	precision, recall, f1, and accuracy
Epochs	50
Batch size	12
Learning rate	0.001

5.5 Environment setup

Experiments conducted in this paper are implemented through utilizing TensorFlow and Keras packages in Python on the Google Collaboratory website.

The input dataset is divided into three subsets: 50% for the training subset, 20% for the validation subset, and 30% for the testing subset.

5.6 Performance evaluation metrics

This paper uses performance evaluation metrics such as accuracy, precision, recall, and F1-score to assess the performance of the DL models. Equations 12 - 15 [62] show the computation of these metrics.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{13}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{14}$$

$$\text{F1} = 2 * \frac{recall * precision}{recall + precision} \tag{15}$$

where TP is an abbreviation for true positives, TN is an abbreviation for true negatives, FP is an abbreviation for false positives, and FN is an abbreviation for false negatives.

5.7 Performance evaluation of different DL Models with the input dataset after applying the preprocessing steps

In this section, the specified DL models are evaluated using the input dataset after applying the preprocessing steps. Table 7 shows the results obtained from each DL model. The green cells show the best performance results.

Table 7. Evaluation results of different DL models using the input dataset after applying the preprocessing steps.

Performance metric	MobileNetV3Large	MobileNetV3Small	ResNet50V2	EfficientNetB0	InceptionV3	NASNetMobile	MobileNetV2
Accuracy	0.9375	0.8646	0.8750	**0.9479**	0.8750	0.8438	0.9271
Precision	0.9365	0.9207	0.8750	**0.9562**	0.8750	0.8710	0.9325
Recall	**0.9167**	0.8438	0.8750	0.9062	0.8750	0.8438	0.8750
F1 score	0.9264	0.8805	0.8750	**0.9303**	0.8750	0.8571	0.9027

5.8 Performance evaluation of different DL Models using the truth NS subset of the input dataset

In this section, the specified DL models are evaluated using the input dataset after applying the preprocessing steps and the NS to extract the truth subset from the input scans using Equation 3. The results obtained from each model are presented in Table 8. The green cells show the best performance results. Figure 2 illustrates samples of the dataset after applying NS to extract the truth subset.

Table 8. Evaluation results of different DL models using the truth NS subset of the input dataset.

Performance metric	MobileNetV3Large	MobileNetV3Small	ResNet50V2	EfficientNetB0	InceptionV3	NASNetMobile	MobileNetV2
Accuracy	0.8542	**0.9375**	0.8854	0.9271	0.8646	0.8438	0.9167
Precision	0.9306	**0.9570**	0.8854	0.9257	0.8646	0.8801	0.9210
Recall	0.8333	**0.9062**	0.8854	**0.9062**	0.8646	0.8438	0.8542
F1 score	0.8792	**0.9307**	0.8854	0.9158	0.8646	0.8612	0.8861

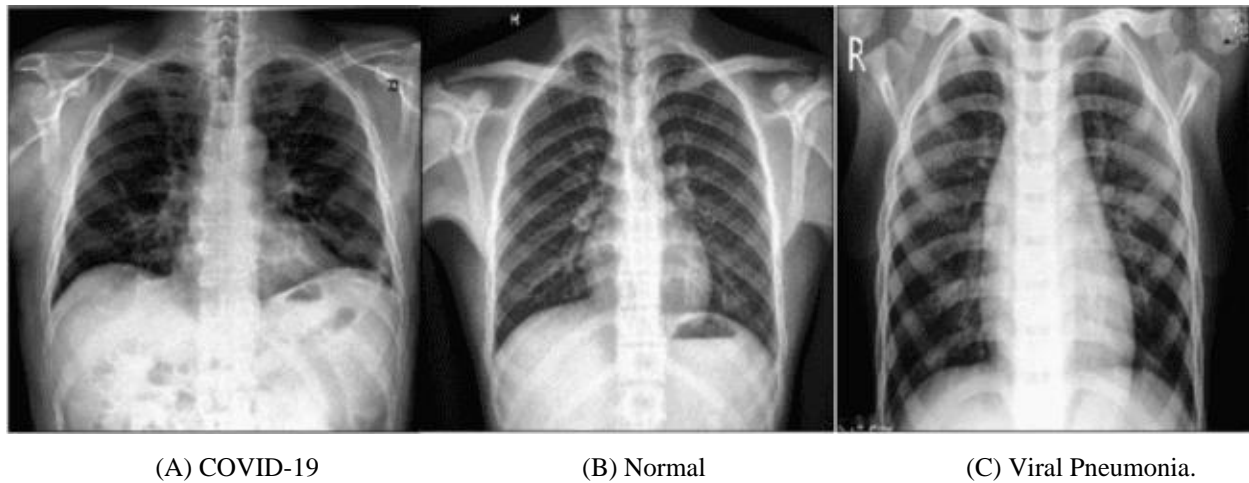(A) COVID-19   (B) Normal   (C) Viral Pneumonia.

Figure 2. Samples of the truth NS images resulting from the input dataset (A) Covid-19, (B) Normal, and (C) Viral Pneumonia.

5.9 Performance evaluation of different DL Models using the falseness NS subset of the input dataset

In this section, the specified DL models are evaluated using the input dataset after applying the preprocessing steps and the NS to extract the falseness subset from input images using Equation 4. The results obtained from each model are presented in Table 9. The green cells show the best performance results. Figure 3 illustrates samples of the dataset after applying NS to extract the falseness subset.

Table 9. Evaluation results of different DL models using the falseness NS subset of the input dataset.

Performance metric	MobileNetV3Large	MobileNetV3Small	ResNet50V2	EfficientNetB0	InceptionV3	NASNetMobile	MobileNetV2
Accuracy	0.9375	0.8958	0.8854	**0.9792**	0.8438	0.8542	0.8854
Precision	0.9462	0.9130	0.8854	**0.9785**	0.8438	0.8501	0.8895
Recall	0.9167	0.8646	0.8854	**0.9583**	0.8438	0.8333	0.8438
F1 score	0.9310	0.8874	0.8854	**0.9683**	0.8437	0.8416	0.8658

(A) COVID-19   (B) Normal   (C)Viral Pneumonia.

Figure 3. Samples of the falseness NS images resulting from the input dataset used (A) Covid-19, (B) Normal, and (C) Viral Pneumonia.

5.10 Performance evaluation of different DL Models using the indeterminacy NS subset of the input dataset

In this section, the specified DL models are evaluated using the input dataset after applying the preprocessing steps and the NS to extract the indeterminacy subset from input images using Equation 5. The results obtained for each model are presented in Table 10. The green cells show the best performance results. Figure 4 illustrates samples of the dataset after applying NS to extract the indeterminacy subset.

Table 10. Evaluation results of different DL models using the indeterminacy NS subset of the input dataset.

Performance metric	MobileNetV3Large	MobileNetV3Small	ResNet50V2	EfficientNetB0	InceptionV3	NASNetMobile	MobileNetV2
Accuracy	**0.9479**	0.8229	0.7708	0.8750	0.7604	0.7708	0.8438
Precision	**0.9666**	0.8455	0.7667	0.8697	0.7848	0.8211	0.8782
Recall	**0.8854**	0.7396	0.7188	0.8333	0.7083	0.6354	0.7604
F1 score	**0.9238**	0.7888	0.7419	0.8508	0.7440	0.7160	0.8149

(A) COVID-19   (B) Normal   (C)Viral Pneumonia.
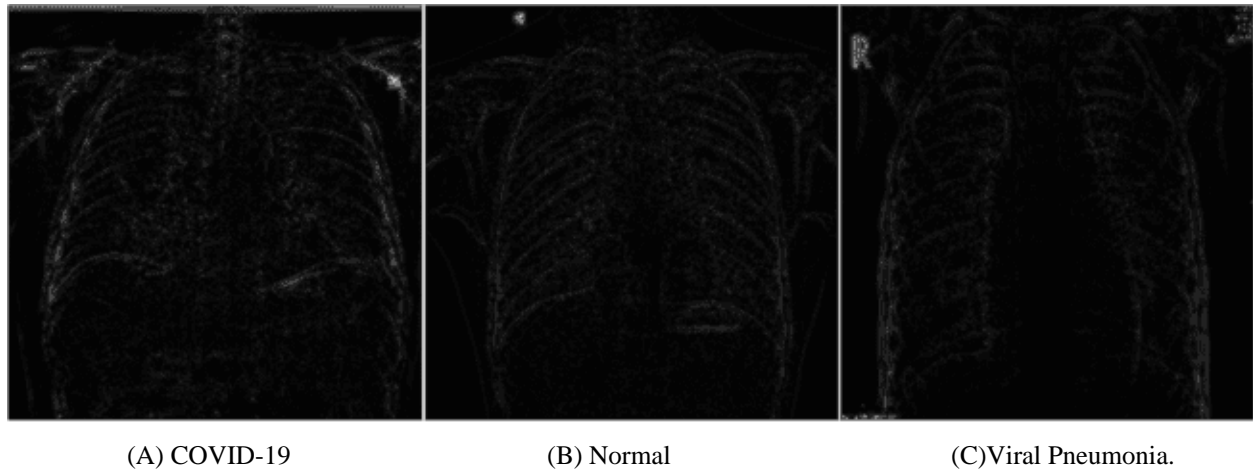
Figure 4. Samples of the indeterminacy NS images from the input dataset (A) Covid-19, (B) Normal, and (C) Viral Pneumonia.

Comparing results shown in Tables 6-8 shows that the results shown in Table 9 are the best NS performance results obtained compared with the results shown in Table 7. Substantial performance improvements are achieved in DL models such as:

- **EfficientNetB0**: the accuracy increases from 0.9479 to 0.9792, the precision increases from 0.9562 to 0.9785, the recall increases from 0.9062 to 0.9583, and the F1 score increases from 0.9303 to 0.9683.
- **MobileNetV3Large**: the accuracy remains at 0.9375, the precision increases from 0.9365 to 0.9462, the recall stays at 0.9167, and the F1 score increases from 0.9264 to 0.9310.
- **ResNet50V2**: the accuracy increases from 0.8750 to 0.8854, the precision remains at 0.8854, the recall increases from 0.8750 to 0.8854, and the F1 score remains at 0.8854.

While the **EfficientNetB0** model shows the most significant improvements, other models such as **MobileNetV3Large** and **ResNet50V2** maintain competitive performance. These improvements are consistent across the accuracy, the precision, the recall, and the F1 score, with no significant degradation in other models. Figure 5 shows the accuracy of different DL models using various subset components of the NS of the input dataset after applying the preprocessing steps. It is shown in this figure that the EfficientNetB0 achieves the most accurate diagnosis results using the falseness NS subset of the input dataset. Consequently, we use the falseness NS subset in the proposed framework.
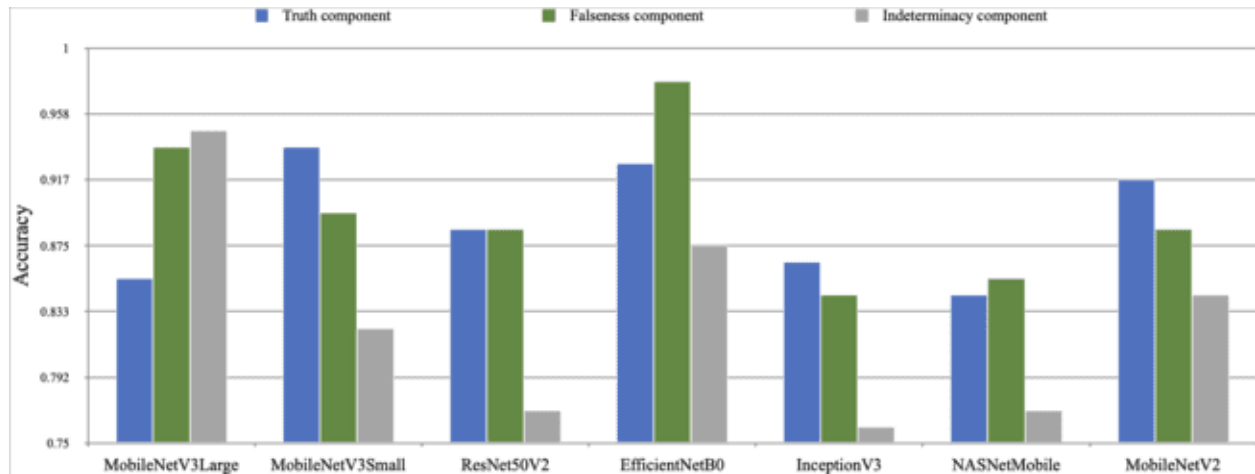
Figure 5. The accuracy results of different DL models using various NS subset components of the input dataset.

5.11   Performance evaluation of different DL Models using data augmentation of the falseness NS subset of the input dataset

After applying data augmentation methods presented in Section 4.3 on the falseness NS subset of the input dataset after applying the preprocessing steps, the size of the subset increases to 2400, and all classes become equally represented. Performance evaluation results of different DL models after applying DA are presented in Table 11. Figure 6 illustrates the accuracy of different DL models before and after Applying DA on the falseness NS subset of the input dataset after applying the preprocessing steps.
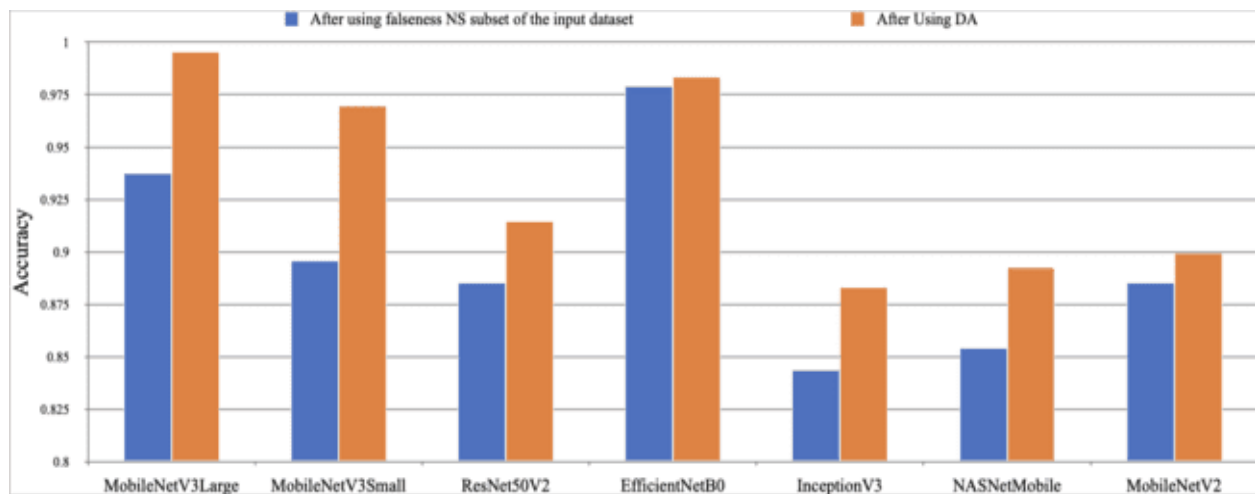
Figure 6. The accuracy of different DL models using the falseness NS subset of the input dataset without and with DA.

Table 11. Results of different DL models after applying DA on the NS Falseness subset of the input dataset.

Performance metric	MobileNetV3Large	MobileNetV3Small	ResNet50V2	EfficientNetB0	InceptionV3	NASNetMobile	MobileNetV2

Accuracy	**0.9953**	0.9696	0.9147	0.9836	0.8832	0.8925	0.8995
Precision	**0.9965**	0.9765	0.9162	0.9871	0.8834	0.8943	0.9132
Recall	**0.9919**	0.9664	0.9151	0.9803	0.8812	0.8931	0.8750
F1 score	**0.9942**	0.9714	0.9156	0.9836	0.8823	0.8937	0.8934

6   Discussion of results

6.1 The impact of using NS on Deep Learning model

The integration of NS into DL models has significantly impacted on their classification performance. This is proved through comparing performance results presented in Table 7 (before using NS) with results presented in Table 9 (after using falseness NS subset of the input dataset). Some models show substantial improvements in their performance. For example, EfficientNetB0 achieves a remarkable increase in accuracy (from 0.9479 to 0.9792), along with in precision (from 0.9562 to 0.9785), recall (from 0.9062 to 0.9583), and F1 score (from 0.9303 to 0.9683). Similarly, ResNet50V2 and MobileNetV3Small also have slight improvements in accuracy and F1 scores, with ResNet50V2's accuracy and F1 score increasing from 0.8750 to 0.8854, and MobileNetV3Small's accuracy score rising from 0.8646 to 0.8958 and F1 score increasing from 0.8805 to 0.8874.

Therefore, the application of NS improves the robustness and reliability of DL models through managing uncertainty and imprecision in data. Figure 7 and Table 12 illustrate the accuracy improvement of different DL due to using falseness NS subset of the input dataset.
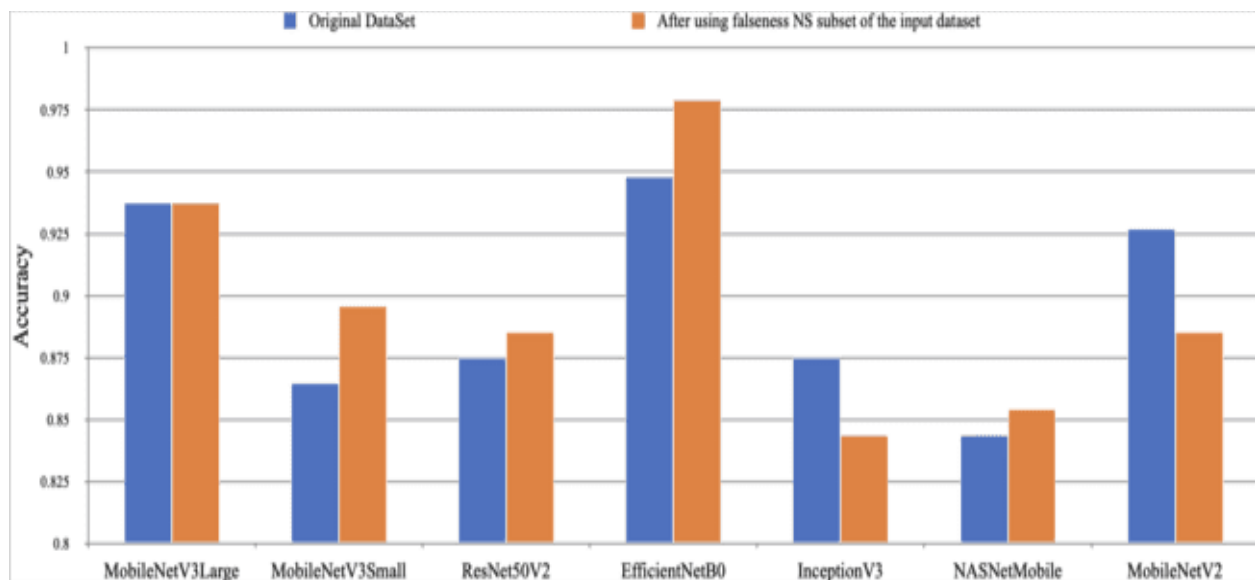

Figure 7. The accuracy of DL models using the input dataset and the falseness NS subset of it.

6.2 The impact of Data Augmentation

Data Augmentation (DA) significantly improves the performance of DL models. This is proved by comparing results shown in Table 9 (before DA) with results shown in Table 11 (after DA). In Table 9, the

EfficientNetB0 achieves the highest accuracy of 0.9792 and an F1 score of 0.9683. However, after applying DA, Table 11 shows that all models exhibit substantial improvements. For example, the MobileNetV3Large achieves an impressive accuracy of 0.9953 and a perfect F1 score of 0.9942, indicating a significant boost in both precision and recall. Similarly, the EfficientNetB0 maintains its strong performance, achieving an accuracy of 0.9836 and an F1 score of 0.9836. Notably, models like the MobileNetV3Small and the ResNet50V2 show large increase in accuracy, precision, recall, and F1 score.

Therefore, DA not only increases the generalization ability of the different DL models but also improves their robustness and reliability leading to better overall performance results across accuracy, precision, recall, and F1 score metrics as shown in Figure 6. Table 13 presents the confusion matrix for each DL model without and with DA using the falseness NS subset of the input dataset after applying the preprocessing steps.

6.3 Proposed framework architecture

Following discussion of results presented in Sections 6.1, and 6.2, an MV3Lung-NS framework architecture, shown in Figure 8, is proposed for the diagnosis of lung infection in CXR scans. This framework architecture is composed of the following steps:
1. Resizing all scans into a fixed size of 224 x 224 pixels.
2. Transforming RGB CXR scans into grayscale scans.
3. Applying NS Filtering to obtain Falseness subset.
4. Applying DA to increase the size of the dataset and balance its classes.
5. Using the MobileNetV3Larg**e** pretrained DL model for scan diagnosis.
6. Using The SGDR algorithm to periodically reduce the learning rate through training, helping the DL model to explore the validation loss landscape, avoid local minima, and find better optima. This results in faster convergence and more accurate results.
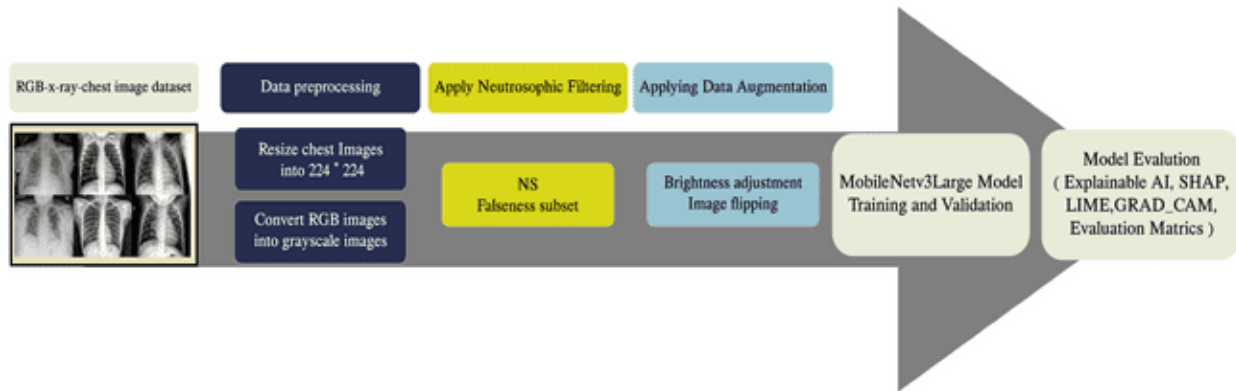

Figure 8. The proposed MV3Lung-NS Framework architecture for lung infection diagnosis using CXR scans.

6.4 Explaining and interpreting results using explainable artificial intelligence

Explainable artificial intelligence (XAI) is necessary for ensuring transparency and understanding of results of artificial intelligence (AI) systems, particularly in the healthcare sector, where AI decisions can significantly affect patient lives. XAI improves the interpretability of AI models, which is vital for establishing trust and helping healthcare professionals to make informed decisions. In the context of lung tumor diagnosis, XAI clarifies the reasoning of AI-generated predictions, which informs clinicians about treatment options.

Several algorithms are proposed for AI decisions interpretability such as Local Interpretable Model-agnostic Explanations (LIME)[27], SHapley Additive exPlanations (SHAP) [26], and Gradient-weighted Class Activation Mapping (Grad-CAM) [28]. LIME identifies which parts of an image influenced a model's decision[27], while SHAP values indicate the importance of various features using principles from game theory [26]. Grad-CAM, often used in neural networks, highlights the specific areas of an image that the model focuses on when making predictions[28]. In lung tumor diagnosis, these algorithms help pinpoint relevant regions in X-ray scans, clarify why certain areas are classified as tumorous, and ensure compliance with regulatory standards regarding transparency. This transparency fosters trust in AI tools, encouraging their responsible use in clinical settings. XAI not only enhances the transparency and reliability of AI systems but also serves a critical function in scenarios such as lung tumor diagnosis. Therefore, by employing LIME, SHAP, and Grad-CAM algorithms, XAI ensures that AI-driven decisions are understandable and reliable, ultimately facilitating effective and informed healthcare delivery.

- **Visualization and interpretation of AI decisions via SHapley Additive exPlanations (SHAP)**

We utilize SHAP algorithm to analyze the predictions made by the proposed framework and to assess the contribution of each feature to the output decision. SHAP values are selected because of their solid foundation in cooperative game theory and their capacity to offer both local and global insights, ensuring equitable and consistent attribution of feature significance. By using SHAP, we can pinpoint the primary factors influencing the AI predictions, providing clarity into the decision-making process at play. This algorithm not only improves transparency but also results in an understanding of the connections between input features and the predicted results of the proposed framework [26].

SHAP provides a detailed explanation of the proposed framework predictions by assigning importance scores to individual pixels in the images. SHAP is effective in identifying significant areas within the images to the output decisions. Pink regions highlight areas that are recognized as important in CXR scans, while blue regions indicate areas that are not important. Additionally, SHAP values illustrate which parts of the image are crucial to the output predictions. Visual representation of these values offers valuable insights into the decision-making process and assists medical professionals in understanding its outcomes. Figure 9 shows the plots visualizing for three samples of CXR scans and the output predictions of the proposed framework using SHAP.

This plot provides a visual representation of how SHAP values contribute to the final classification decision for a specific image. It clearly indicates how certain features influence the prediction towards one class over the others. For instance, if a scan is diagnosed as COVID-19, the plot shows that features consistent with COVID characteristics—such as specific patterns observed in lung imaging—have high positive SHAP values.

In a SHAP summary plot, features are ranked by their importance based on absolute SHAP values, with each point representing the SHAP value for a specific feature in an individual instance. If a feature, like a particular lung pattern, has many red points (indicating high feature values), it suggests that this feature plays a significant role in predicting a specific class. Conversely, features associated with other classes show negative or neutral contributions. This explanation produces interpretation of the diagnosis process of the DL model.
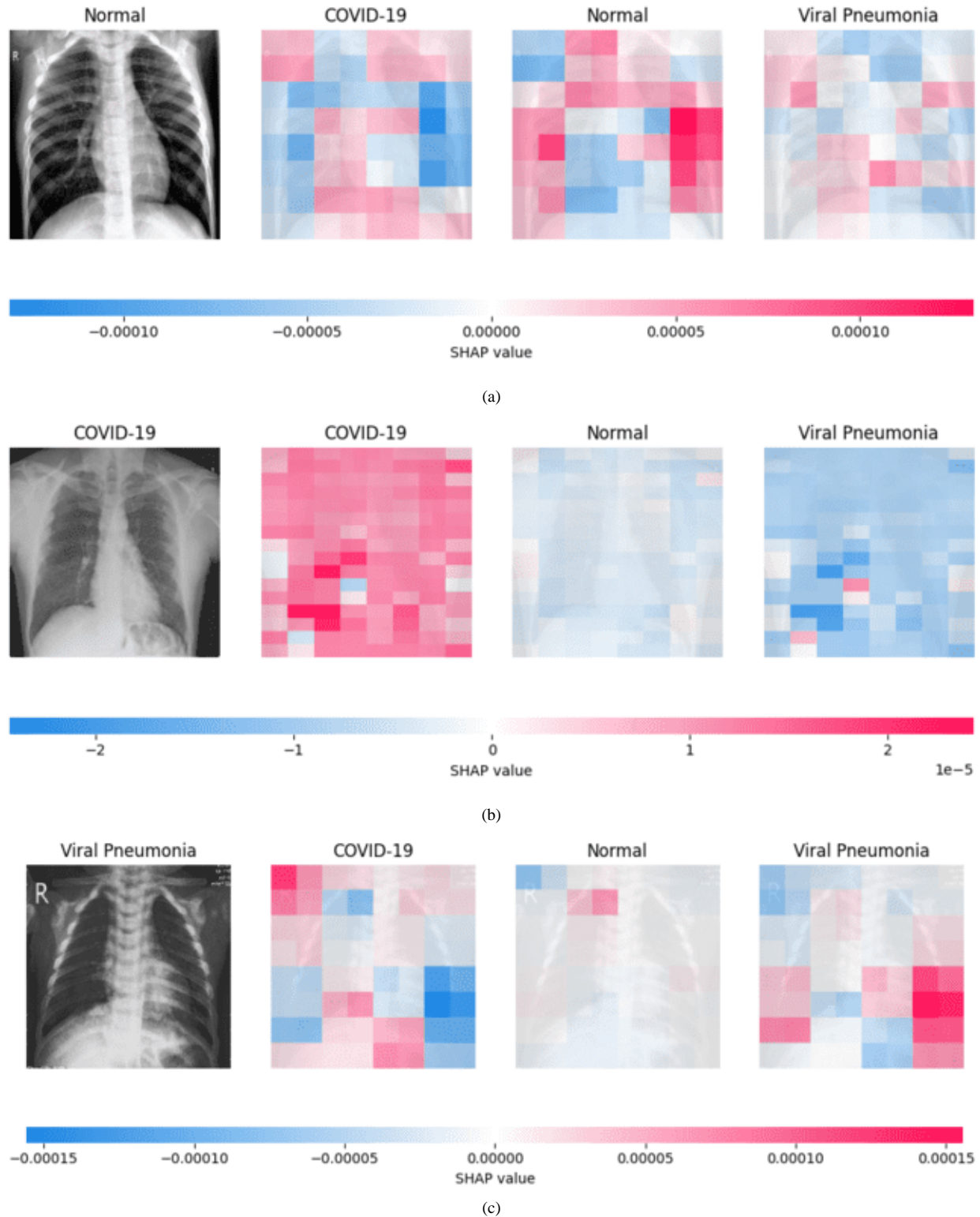
Figure 9. SHAP plots visualizing the neural network's distinguish process for input dataset (A) Normal image (B) Covid-19 image and (C) Viral Pneumonia Image.

- **Visualization and interpretation of AI decisions via Local Interpretable Model-agnostic Explanations (LIME)**

LIME algorithm is introduced by Ribeiro et al [27]. It helps users understand the contribution of individual input features to an AI model's output, making it especially valuable in high-stakes fields like healthcare and finance. LIME algorithm is model agnostic i.e., it can be applied to a wide variety of algorithms, including deep neural networks and ensemble methods. However, it does have limitations including sensitivity to parameter choices and potential instability in the explanations provided [63]. Despite these drawbacks, LIME remains a fundamental tool in the field of XAI, helping human in interpreting complex DL model's output. In CXR scan classification, LIME identifies important regions by altering the input image, training a simple interpretable model, and highlighting key areas that influence predictions. For instance, if a model classifies a CXR scan as "COVID-19 positive," LIME may emphasize specific lung regions, such as those exhibiting ground-glass opacities, to illustrate the reasoning behind the model's decision. This approach is particularly beneficial in medical imaging, as understanding why a model predicts "COVID-19," "pneumonia," or "normal" is crucial for clinicians to trust and validate the model's outcomes. Figure 10 show LIME visualizations that elucidate the neural network's decision-making process for the input images, providing insights into the model's interpretability.
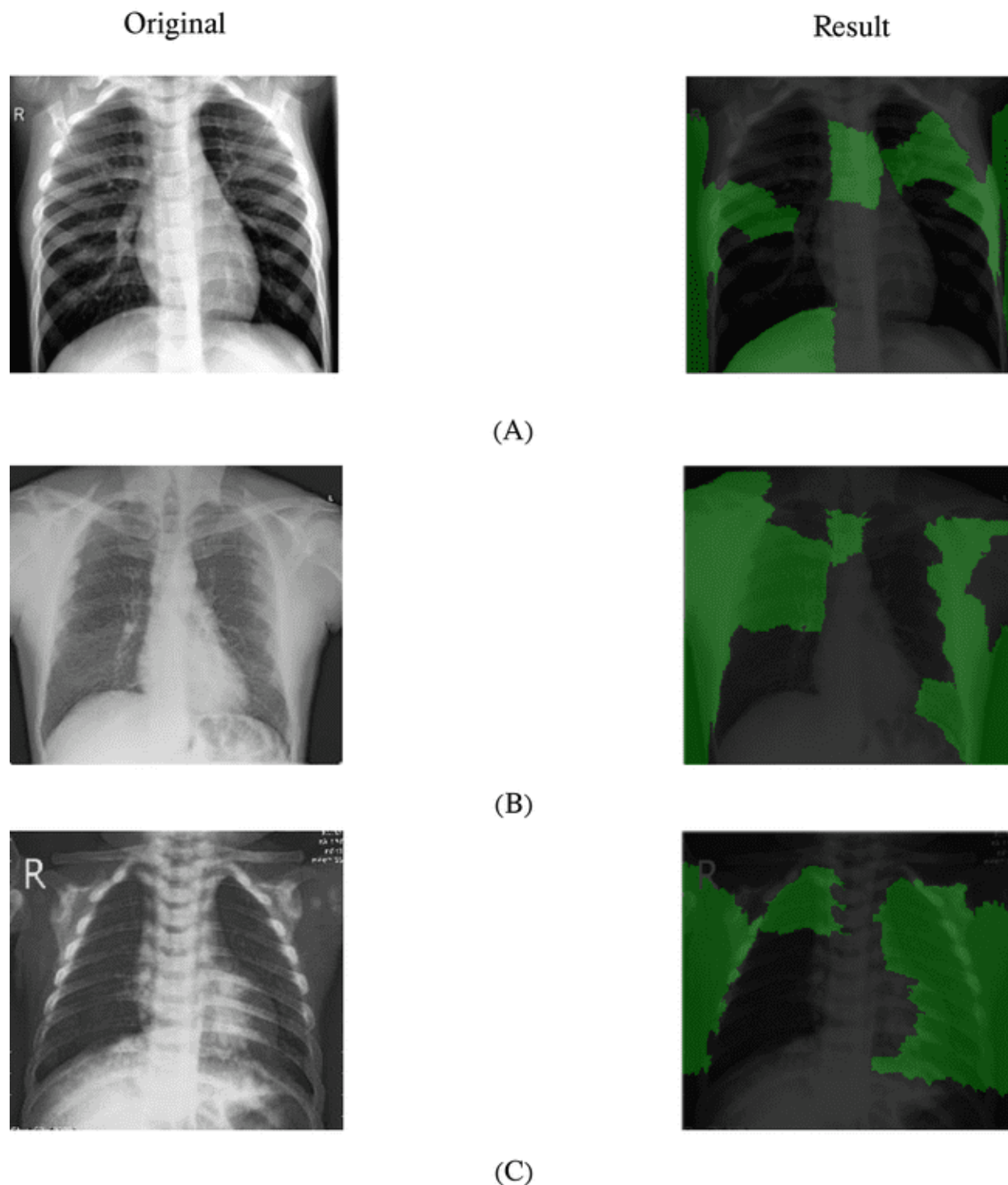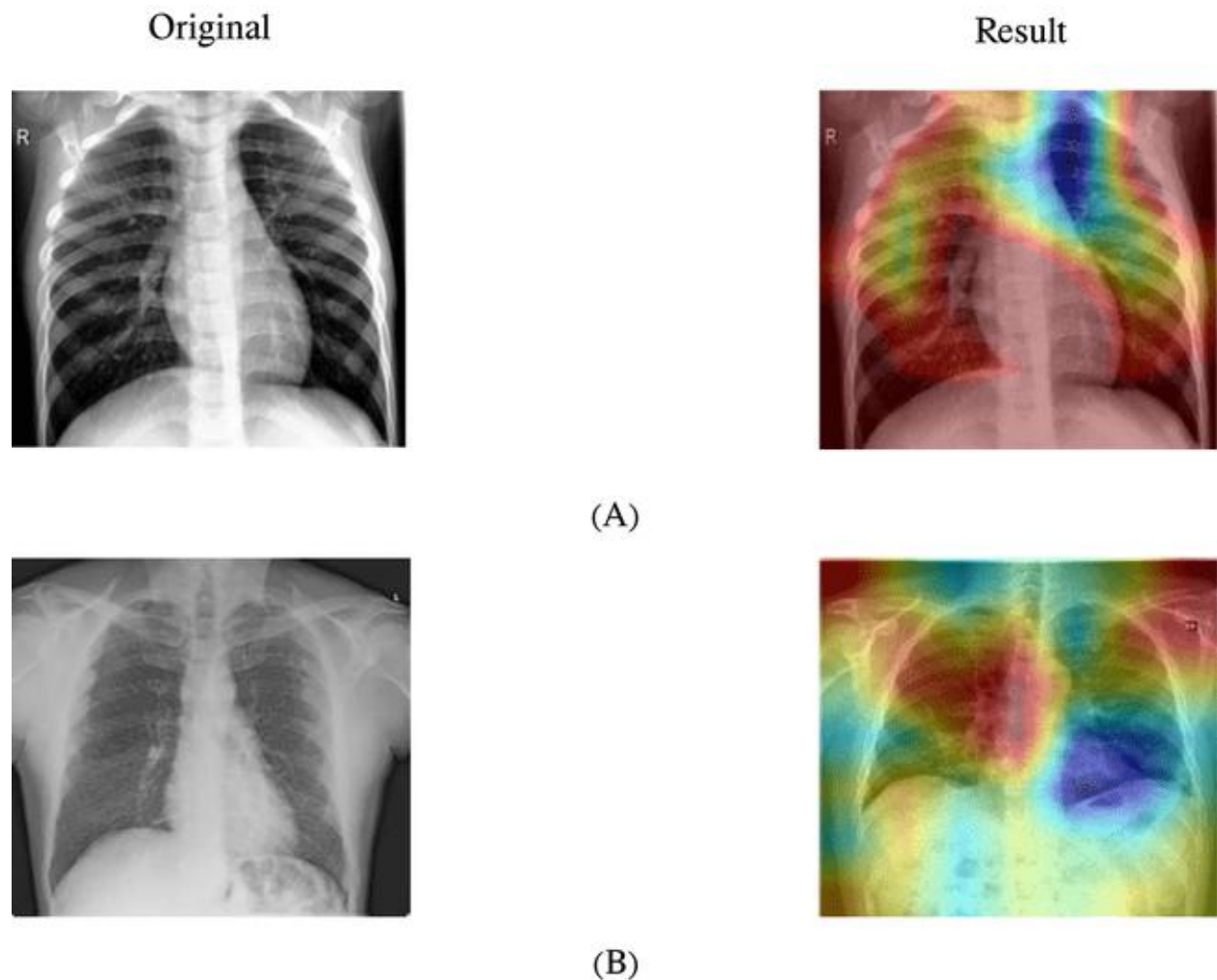
Figure 10. LIME visualizations illustrating the neural network's decision-making process for input scans: (A) Normal (B) Covid-19 (C) Viral Pneumonia.

- **Visualization and interpretation of AI decisions via Gradient-weighted Class Activation Mapping (Grad-CAM) algorithm**

Grad-CAM algorithm is a visualization method used to interpret and explain the decision-making processes of convolutional neural networks (CNNs). It increases the gradients of a target class that flow into the final convolutional layer to generate a coarse localization map. This map shows the segments of an input scan that have large contributions to the model's prediction. Grad-CAM is class-specific i.e., it can highlight areas relevant to a particular class even in multi-class classification tasks. The resulting heatmap uses a color gradient, typically ranging from cool colors (like blue) to warm colors (like red), to indicate the importance of various regions. Warmer colors signify higher relevance to the model's decision. This technique is invaluable for enhancing model interpretability, debugging and validating whether the model focuses on meaningful features. It is particularly important in applications such as medical imaging and autonomous driving. Grad-CAM's ability to provide visual explanations without requiring changes to the model architecture or retraining makes it a widely adopted tool in deep learning research [28, 64, 65]. Figure 11 displays Grad-CAM visualizations, highlighting the neural network's focus areas and decision-making process for the input images.
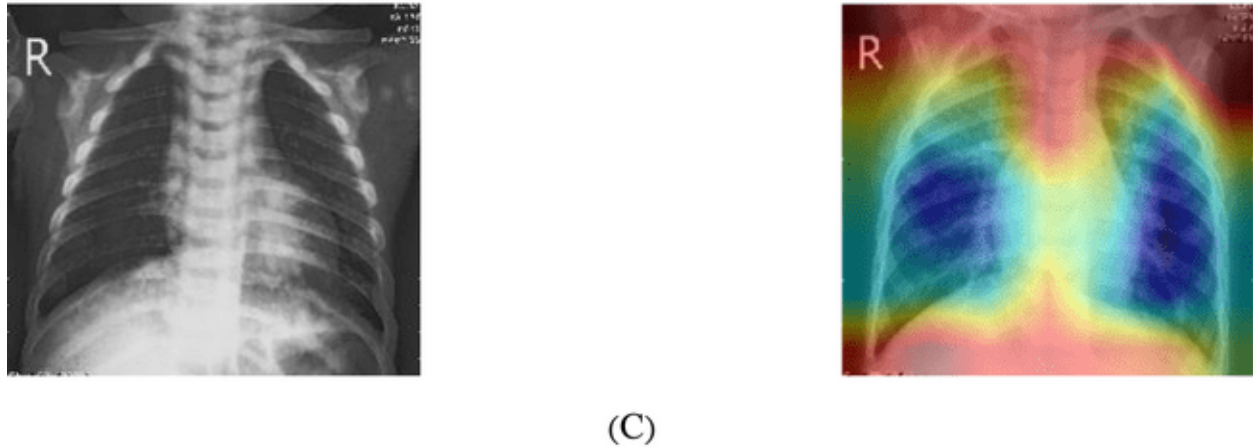

(A)

(B)

(C)

Figure 11. Grad-CAM visualizations highlighting the neural network's decision-making process for input scans (A) Normal (B) Covid-19 (C) Viral Pneumonia.

6.5 Advantages of the MV3Lung-NS Proposed Framework

- **Accurate and reliable performance:**
  The MV3Lung-NS framework achieves impressive diagnostic performance, with an accuracy of 99.53%. This surpasses the results of other deep learning models listed in Table 14 and addresses the limitations outlined in Table 15. Additionally, it boasts a precision of 99.65%, a recall of 99.19%, and an F1-score of 99.42%. These results underscore its robustness and reliability across various evaluation criteria.

- **Capability of analyzing uncertainty in medical image analysis:**
  Using the NS allows the MV3Lung-NS Proposed Framework to efficiently handle uncertainty, indeterminacy, and imprecision in medical images. By explicitly modeling truth, falsehood, and indeterminacy, NS improves feature representation and enhances performance robustness.

- **Effective Handling of Class Imbalance:**
  Using DA significantly addresses class imbalance in medical imaging datasets by creating diverse samples for under-represented classes, thereby ensuring balanced and effective model training.

- **Visualization and interpretation of the diagnosis results:**
  Utilizing XAI methods in the MV3Lung-NS Proposed Framework improves transparency and builds trust among clinicians by elucidating how the model makes decisions, an essential factor for effective clinical application.

7 Conclusion and future work

The integration of Neutrosophic Sets and Data Augmentation has proven to be highly effective in enhancing the performance of DL models for lung-infection diagnosis using CXR scans. NS improve model robustness by addressing data uncertainty while DA significantly boosted generalization and accuracy across multiple architectures. The proposed MV3Lung-NS framework, which combines NS, DA, and the MobileNetV3Large model, emerged as an efficient tool for the diagnosis of lung infections. The MV3Lung-NS framework achieves new state-of-the-art diagnosis performance metrics' results that are an accuracy of

99.53%, a precision of 99.65%, a recall of 99.19%, and an F1-score of 99.42%. Furthermore, the application of XAI methods, such as SHAP, LIME, and Grad-CAM, provides critical insights into model predictions, ensuring transparency and interpretability of the obtained decision results. These advancements not only improve model performance but also foster trust and reliability in the proposed framework, paving the way for its responsible adoption in clinical settings. Future work could explore extending this framework to other medical imaging domains and further refining XAI methods for enhanced interpretability. In addition, more advanced or domain-specific DA techniques could be investigated to improve the quality of the input dataset.

Table 12. Confusion matrix of DL models using the input dataset and the falseness NS subset.
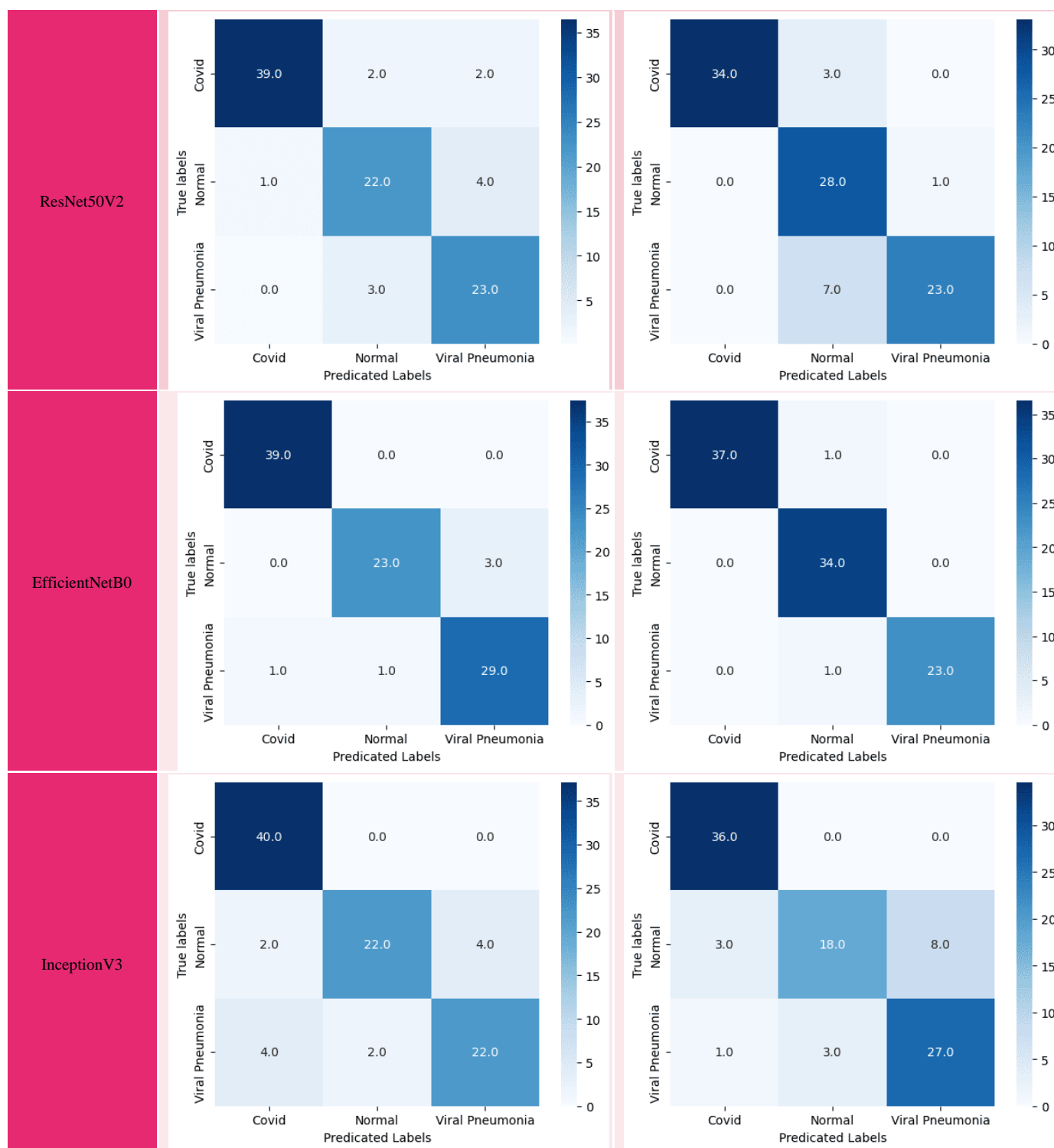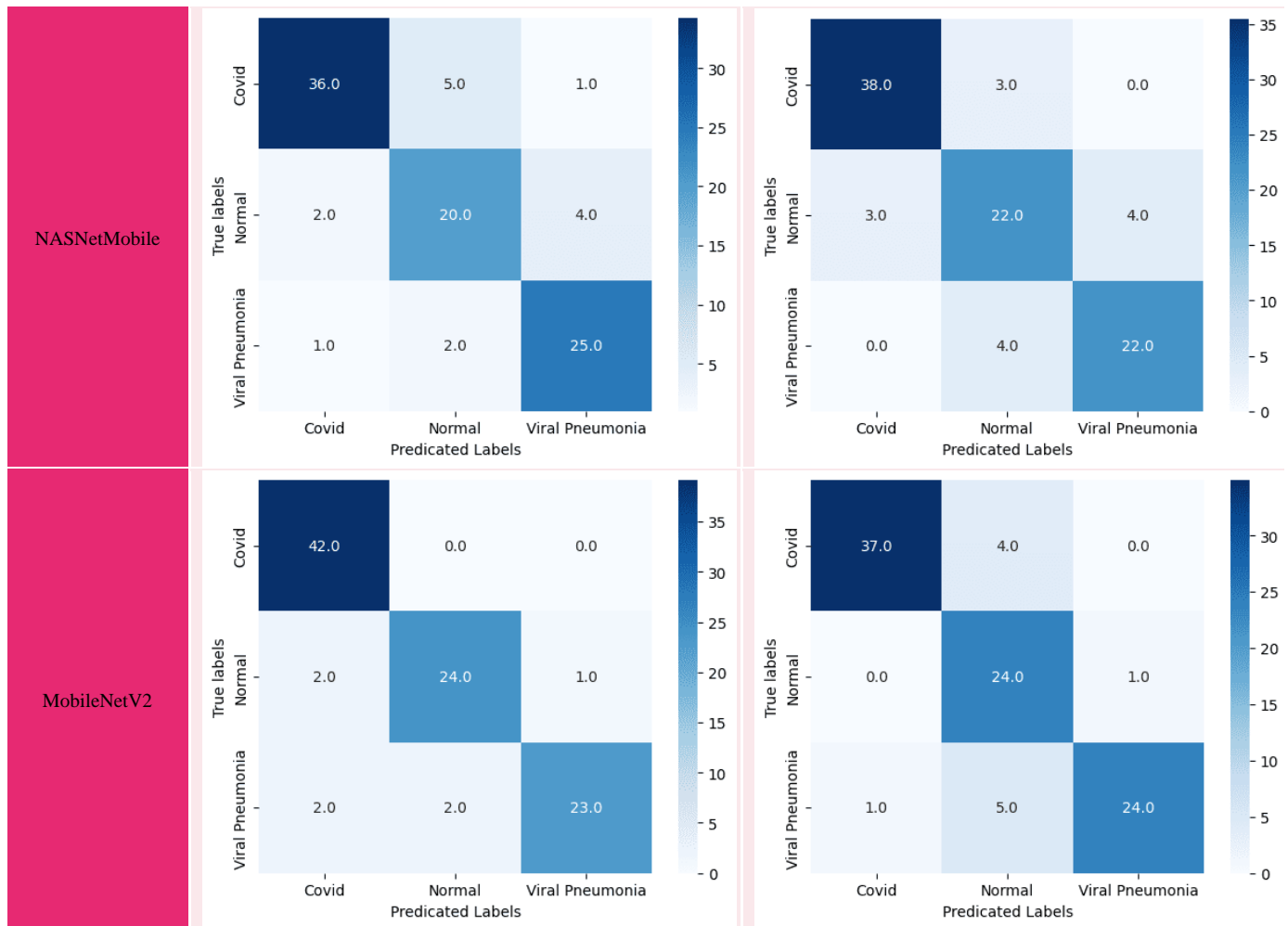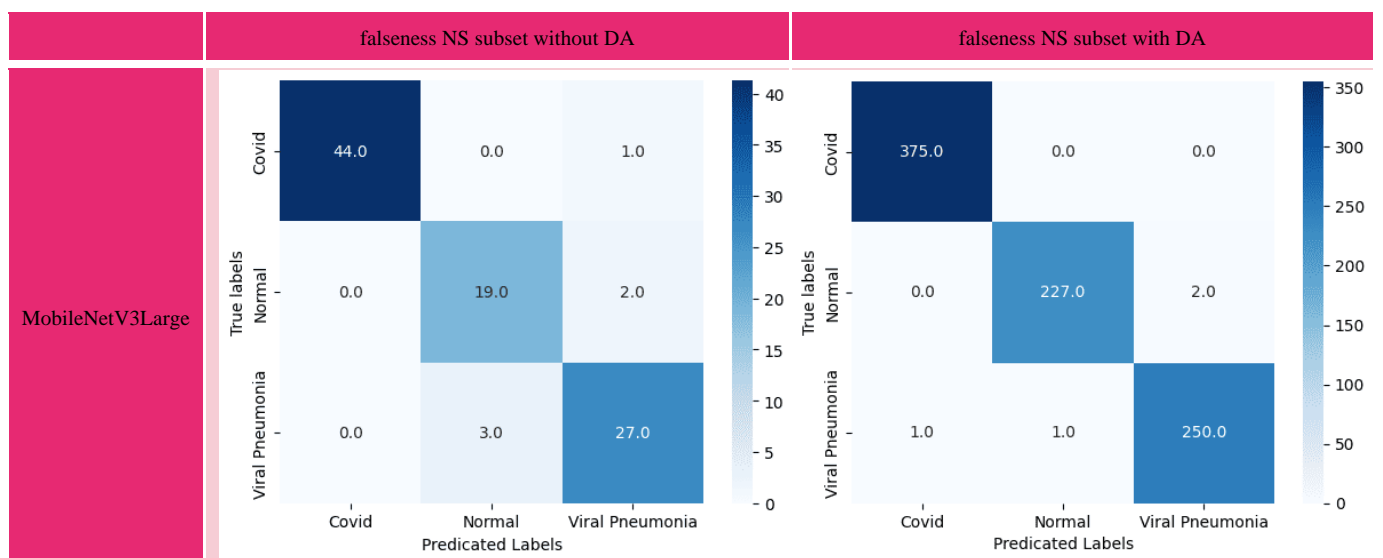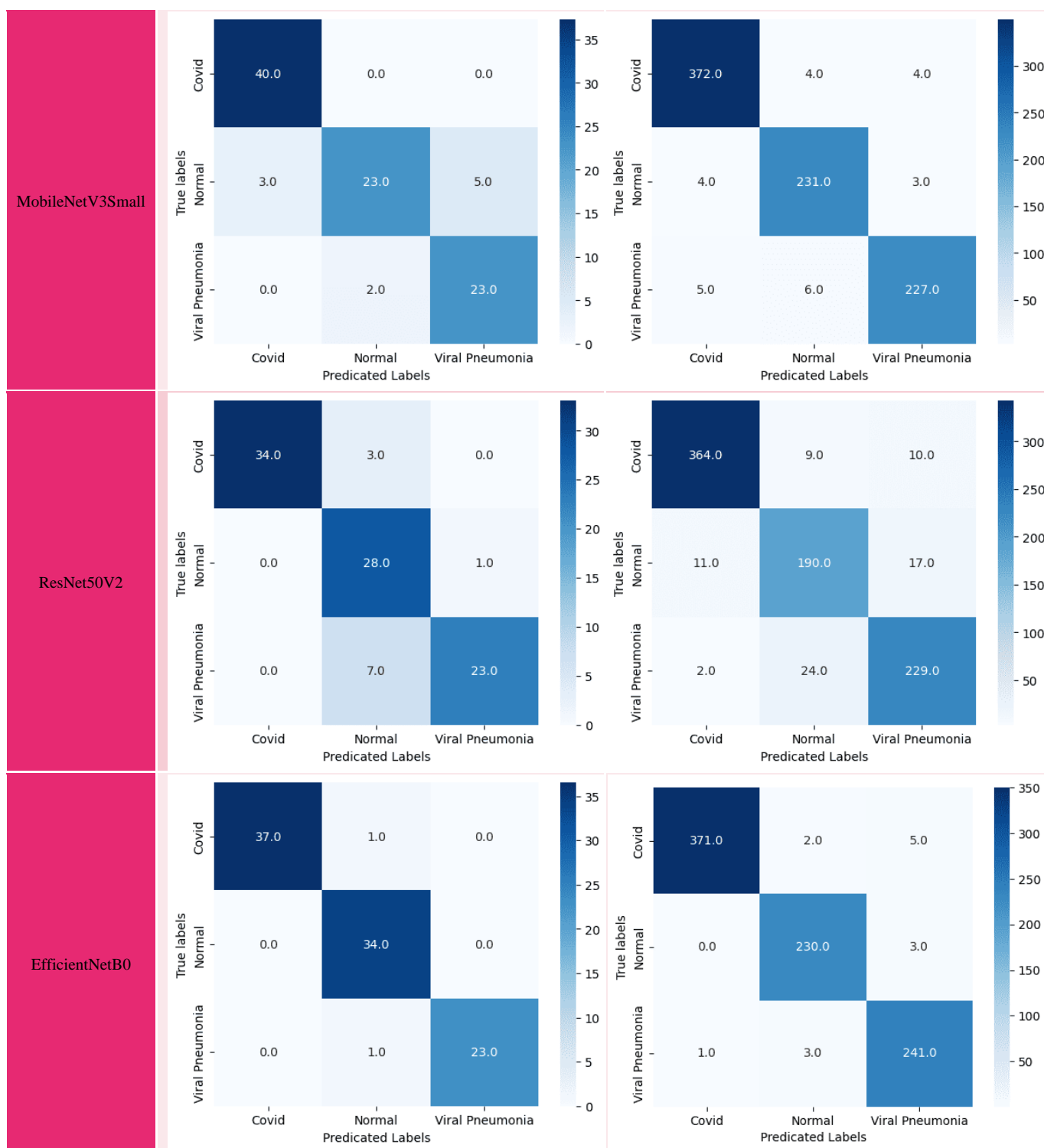
Table 13. Confusion matrix of different DL models using the falseness NS subset of the input dataset without and with DA.
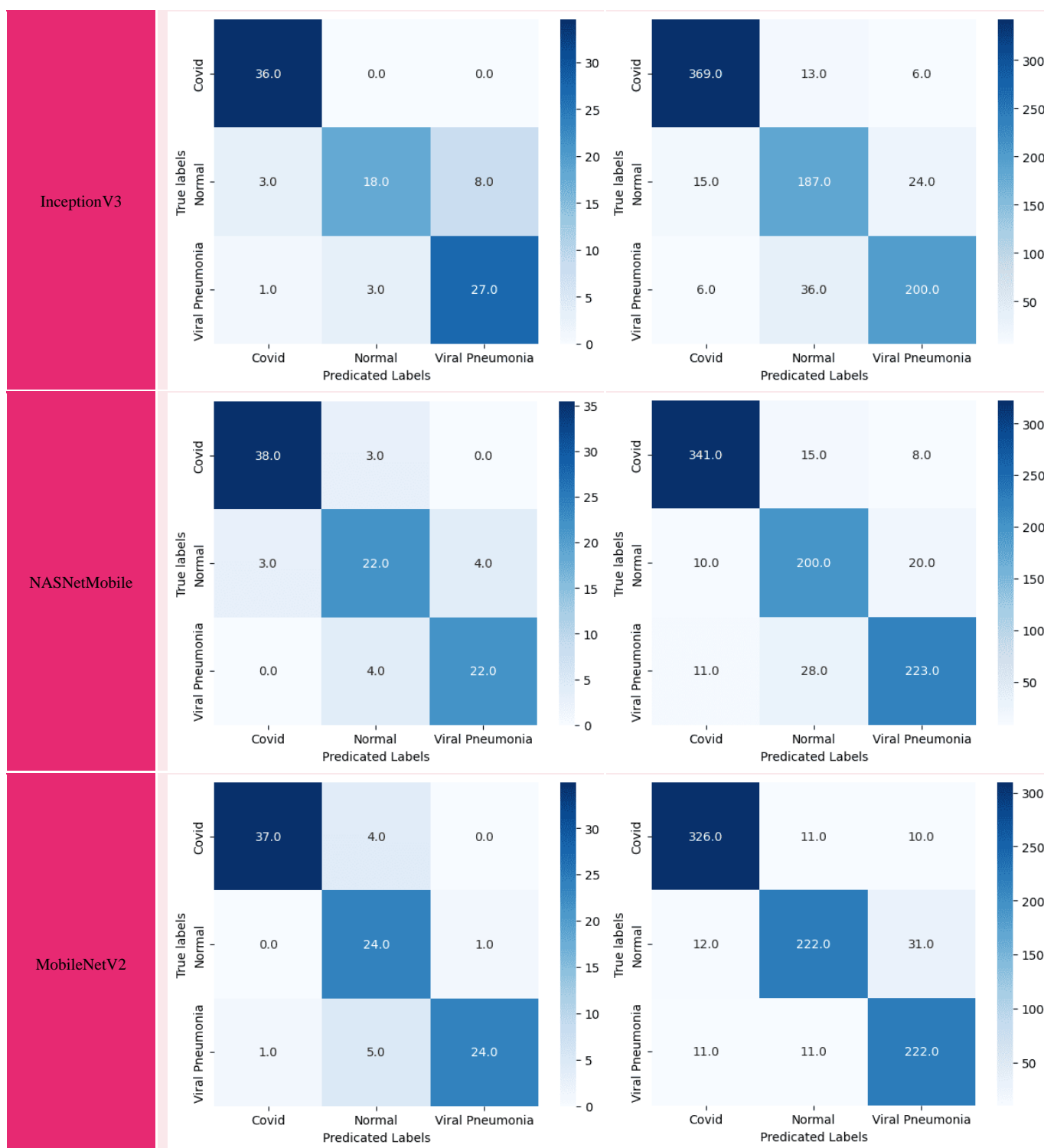
Table 14. Comparing performance of the proposed MV3Lung-NS framework with other DL models in the literature.

Reference	Model	Dataset	N. Classes	Accuracy	Precision	Recall	F1
[66]	FPS Optimization	Covid-19 Image Dataset	3	0.9615	-	-	-
[67]	Deep CNN (VGG-16)	Covid-19 Image Dataset	3	0.957	0.889	0.888	-
[68]	Trained Output-based Transfer Learning	Covid-19 Image Dataset	3	0.966	0.9892	0.9388	0.9634
[69]	ResNet101	Covid-19 Image Dataset	3	0.963	-	-	-
Proposed framework	**MV3Lung-NS**	**Covid-19 Image Dataset**	3	**0.9953**	**0.9965**	**0.9919**	**0.9942**

Table 15. Limitation of DL models in the literature

Reference	Model	Limitations
[66]	FPS Optimization	There is no pre-processing step for denoising images in the dataset.
[67]	Deep CNN (VGG-16)	The model has several limitations, including compatibility issues with image and numerical datasets, sensitivity to low-quality images, imbalanced data classes, and dataset size.
[68]	Trained Output-based Transfer Learning	The introduced model handles datasets without addressing unbalanced data classes issues, which results in low performance results.
[69]	ResNet101	The introduced model handles datasets without addressing unbalanced data classes issues, which results in low performance results. In addition, there is no pre-processing step for denoising images in the dataset.

References

1. Troeger C, Forouzanfar M, Rao PC, et al (2017) Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory tract infections in 195 countries: a systematic analysis for the Global Burden of Disease Study 2015. Lancet Infect Dis 17:1133–1161

2. Febbo J, Dako F (2024) Pulmonary Infection. Clin Chest Med 45:373–382

3. Wang C, Horby PW, Hayden FG, Gao GF (2020) A novel coronavirus outbreak of global health concern. The lancet 395:470–473

4. Karunasagar I, Karunasagar I (2020) Ongoing COVID-19 global crisis and scientific challenges. Journal of Health and Allied Sciences NU 10:1–2

5. Organization WH (2020) Coronavirus disease (COVID-19): situation report, 142. In: Coronavirus disease (COVID-19): situation report, 142

6. Bhattacharya S, Reddy Maddikunta PK, Pham Q-V, et al (2021) Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey. Sustain Cities Soc 65:102589. https://doi.org/https://doi.org/10.1016/j.scs.2020.102589

7. Jia S, Wang Y, Wang W, et al (2022) Value of medical imaging artificial intelligence in the diagnosis and treatment of new coronavirus pneumonia. Expert Syst 39:e12740

8. Chen L, Li S, Bai Q, et al (2021) Review of Image Classification Algorithms Based on Convolutional Neural Networks. Remote Sens (Basel) 13:4712

9. Rubin GD, Ryerson CJ, Haramati LB, et al (2020) The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the Fleischner Society. Radiology 296:172–180

10. Cohen JP, Morrison P, Dao L (2020) COVID-19 image data collection. arXiv preprint arXiv:200311597

11. Benmalek E, Elmhamdi J, Jilbab A (2021) Comparing CT scan and chest X-ray imaging for COVID-19 diagnosis. Biomedical Engineering Advances 1:100003. https://doi.org/https://doi.org/10.1016/j.bea.2021.100003

12. Göreke V (2023) A novel method based on Wiener filter for denoising Poisson noise from medical X-Ray images. Biomed Signal Process Control 79:104031. https://doi.org/https://doi.org/10.1016/j.bspc.2022.104031

13. Chandra TB, Verma K (2020) Analysis of quantum noise-reducing filters on chest X-ray images: A review. Measurement 153:107426. https://doi.org/https://doi.org/10.1016/j.measurement.2019.107426

14. Thakur K V, Damodare OH, Sapkal AM (2016) Poisson Noise Reducing Bilateral Filter. Procedia Comput Sci 79:861–865. https://doi.org/https://doi.org/10.1016/j.procs.2016.03.087

15. Kaur G, Sharma N, Chauhan R, et al (2023) Eye Disease Classification Using ResNet-18 Deep Learning Architecture. In: 2023 2nd International Conference on Futuristic Technologies (INCOFT). pp 1–5

16. Mathew A, Amudha P, Sivakumari S (2021) Deep Learning Techniques: An Overview. In: Hassanien AE, Bhatnagar R, Darwish A (eds) Advanced Machine Learning Technologies and Applications. Springer Singapore, Singapore, pp 599–608

17. Al-Ameen Z, Sulong G (2016) Prevalent degradations and processing challenges of computed tomography medical images: A compendious analysis. International Journal of Grid and Distributed Computing 9:107–118

18. Deheyab AOA, Alwan MH, Rezzaqe I khalid A, et al (2022) An overview of challenges in medical image processing. In: Proceedings of the 6th international conference on future networks & distributed systems. pp 511–516

19. Wang X, Dong H (2023) Click-through rate prediction and uncertainty quantification based on bayesian deep learning. Entropy 25:406

20. Howard A, Sandler M, Chu G, et al (2019) Searching for mobilenetv3. In: Proceedings of the IEEE/CVF international conference on computer vision. pp 1314–1324

21. He K, Zhang X, Ren S, Sun J (2016) Identity Mappings in Deep Residual Networks. In: Leibe B, Matas J, Sebe N, Welling M (eds) Computer Vision – ECCV 2016. Springer International Publishing, Cham, pp 630–645

22. Tan M, Le Q (2019) Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. PMLR, pp 6105–6114

23. Szegedy C, Vanhoucke V, Ioffe S, et al (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 2818–2826

24. Zoph B, Vasudevan V, Shlens J, Le Q V (2018) Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 8697–8710

25. Sandler M, Howard A, Zhu M, et al (2018) Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 4510–4520

26. Lundberg S (2017) A unified approach to interpreting model predictions. arXiv preprint arXiv:170507874

27. Ribeiro MT, Singh S, Guestrin C (2016) " Why should i trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. pp 1135–1144

28. Selvaraju RR, Cogswell M, Das A, et al (2017) Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. pp 618–626

29. Nguyen GN, Son LH, Ashour AS, Dey N (2019) A survey of the state-of-the-arts on neutrosophic sets in biomedical diagnoses. International Journal of Machine Learning and Cybernetics 10:1–13. https://doi.org/10.1007/s13042-017-0691-7

30. Atanassov KT, Atanassov KT (1999) Intuitionistic fuzzy sets. Springer

31. Turksen IB (1986) Interval valued fuzzy sets based on normal forms. Fuzzy Sets Syst 20:191–210

32. Atanassov KT (1989) More on intuitionistic fuzzy sets. Fuzzy Sets Syst 33:37–45

33. Guo Y, Cheng H-D (2009) New neutrosophic approach to image segmentation. Pattern Recognit 42:587–595

34. Guo Y, Cheng HD, Zhang Y (2009) A new neutrosophic approach to image denoising. New Mathematics and Natural Computation 5:653–662

35. Jennifer JS, Sharmila TS (2023) A neutrosophic set approach on chest X-rays for automatic lung infection detection. Information Technology and Control 52:37–52

36. Hemdan EED, Shouman MA, Karar ME (2020) A framework of deep learning classifiers to diagnose COVID-19 in X-Ray images. arXiv preprint arXiv:200311055

37. Mohan J, Krishnaveni V, Guo Y (2013) A new neutrosophic approach of Wiener filtering for MRI denoising. Measurement Science Review 13:177–186

38.   Koundal D, Sharma B (2019) Challenges and future directions in neutrosophic set-based medical image analysis. In: Neutrosophic Set in Medical Image Analysis. Elsevier, pp 313–343

39.   Ravi R, Al-Attar B, Qudr L, et al (2024) Efforts of Neutrosophic Logic in Medical Image Processing and Analysis. International Journal of Neutrosophic Science 24:376–388. https://doi.org/10.54216/IJNS.240428

40.   Chai JS, Selvachandran G, Smarandache F, et al (2021) New similarity measures for single-valued neutrosophic sets with applications in pattern recognition and medical diagnosis problems. Complex & Intelligent Systems 7:703–723

41.   Liu X, Pedersen M, Wang R (2022) Survey of natural image enhancement techniques: Classification, evaluation, challenges, and perspectives. Digit Signal Process 127:103547. https://doi.org/https://doi.org/10.1016/j.dsp.2022.103547

42.   Wady SH, Yousif RZ, Hasan HR (2020) A Novel Intelligent System for Brain Tumor Diagnosis Based on a Composite Neutrosophic-Slantlet Transform Domain for Statistical Texture Feature Extraction. Biomed Res Int 2020:8125392. https://doi.org/https://doi.org/10.1155/2020/8125392

43.   Khalifa NEM, Smarandache F, Manogaran G, Loey M (2024) A Study of the Neutrosophic Set Significance on Deep Transfer Learning Models: an Experimental Case on a Limited COVID-19 Chest X-ray Dataset. Cognit Comput 16:1602–1611. https://doi.org/10.1007/s12559-020-09802-9

44.   Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25:

45.   Szegedy C, Liu W, Jia Y, et al (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 1–9

46.   He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 770–778

47.   Cai G, Guo Y, Chen W, et al (2019) Neutrosophic set-based deep learning in mammogram analysis. In: Neutrosophic Set in Medical Image Analysis. Elsevier, pp 287–310

48.   Guo Y, Ashour AS (2019) Neutrosophic multiple deep convolutional neural network for skin dermoscopic image classification. In: Neutrosophic set in medical image analysis. Elsevier, pp 269–285

49.   Özyurt F, Sert E, Avci E, Dogantekin E (2019) Brain tumor detection based on Convolutional Neural Network with neutrosophic expert maximum fuzzy sure entropy. Measurement 147:106830

50.   Yasser Ibrahim and Abd El-Khalek AA and TA and A-EM-E and SAA and KF (2022) A Hybrid Automated Intelligent COVID-19 Classification System Based on Neutrosophic Logic and Machine Learning Techniques Using Chest X-Ray Images. In: Hassanien Aboul-Ella and Elghamrawy SM and ZI (ed) Advances in Data Science and Intelligent Data Communication Technologies for COVID-19: Innovative Solutions Against COVID-19. Springer International Publishing, Cham, pp 119–137

51.   Hu K, Zhao L, Feng S, et al (2022) Colorectal polyp region extraction using saliency detection network with neutrosophic enhancement. Comput Biol Med 147:105760

52.   Guo Y, Shahin AI, Garg H (2024) An indeterminacy fusion of encoder-decoder network based on neutrosophic set for white blood cells segmentation. Expert Syst Appl 246:123156. https://doi.org/https://doi.org/10.1016/j.eswa.2024.123156

53.   Abdullah W (2025) Enhancing Medical X-Ray Image Classification with Neutrosophic Set Theory and Advanced Deep Learning Models. Neutrosophic Sets and Systems 81:675–698

54.   Elbehiery H, Eassa M, Abdelhafeez A, et al (2025) Advanced Machine Learning Approaches for Breast Cancer Detection with Neutrosophic Sets. Neutrosophic Sets and Systems 81:273–284

55.   Bengio Y, Goodfellow I, Courville A (2017) Deep learning. MIT press Cambridge, MA, USA

56.   Shorten C, Khoshgoftaar TM (2019) A survey on Image Data Augmentation for Deep Learning. J Big Data 6:60. https://doi.org/10.1186/s40537-019-0197-0

57.   Covid-19 Image Dataset https://www.kaggle.com/datasets/pranavraikokte/covid19-image-dataset

58.   Géron A (2022) Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. " O'Reilly Media, Inc."

59.   Rodriguez JD, Perez A, Lozano JA (2010) Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. IEEE Trans Pattern Anal Mach Intell 32:569–575. https://doi.org/10.1109/TPAMI.2009.187

60.   Loshchilov I, Hutter F (2016) Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:160803983

61.   Park A, Chute C, Rajpurkar P, et al (2019) Deep Learning-Assisted Diagnosis of Cerebral Aneurysms Using the HeadXNet Model. JAMA Netw Open 2:e195600. https://doi.org/10.1001/jamanetworkopen.2019.5600

62.   Kuhn M, Johnson K (2013) Applied predictive modeling. Springer

63.   Alvarez-Melis D, Jaakkola TS (2018) On the robustness of interpretability methods. arXiv preprint arXiv:180608049

64.   Zhou B, Khosla A, Lapedriza A, et al (2016) Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 2921–2929

65.   Rajpurkar P (2017) CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. ArXiv abs/1711 5225:

66.   Kandati DR, Gadekallu TR (2023) Federated Learning Approach for Early Detection of Chest Lesion Caused by COVID-19 Infection Using Particle Swarm Optimization. 12:710

67.   Ghnemat R, Alodibat S, Abu Al-Haija Q (2023) Explainable Artificial Intelligence (XAI) for deep learning based medical imaging classification. J Imaging 9:177

68.   Kumar S, Mallik A (2023) COVID-19 Detection from Chest X-rays Using Trained Output Based Transfer Learning Approach. Neural Process Lett 55:2405–2428. https://doi.org/10.1007/s11063-022-11060-9

69.   Yenikaya MA, Kerse G, Oktaysoy O (2024) Artificial intelligence in the healthcare sector: comparison of deep learning networks using chest X-ray images. Front Public Health 12: