Utilizing Deep Learning Algorithms for the Prompt Identification of Chronic Obstructive Pulmonary Disease

Carlos Medina-Ramos
Faculty of Electrical and Electronic Engineering, Universidad Nacionalde Ingenieria, Lima, Peru
ccmedina@uni.edu.pe

Nilton Sare-Vargas
Faculty of Electrical and Electronic Engineering, Universidad Nacional de Ingenieria, Lima, Peru
nilton.sare.v@uni.pe

Warren Reategui-Romero
Faculty of Chemical and Textile Engineering, Universidad Nacional de Ingenieria, Lima, Peru
wreategui@uni.edu.pe

Karin Paucar-Cuba
Faculty of Chemical and Textile Engineering, Universidad Nacional de Ingenieria, Lima, Peru
kpaucar@uni.edu.pe

Daniel Carbonel-Olazabal
Faculty of Electrical and Electronic Engineering, Universidad Nacional de Ingenieria, Lima, Peru
d_carbonel@uni.edu.pe (corresponding author)

Judith Betetta-Gomez
Faculty of Electrical and Electronic Engineering, Universidad Nacional de Ingenieria, Lima, Peru
jbetetta@uni.edu.pe

Received: 27 February 2025 | Revised: 12 April 2025 and 24 April 2025 | Accepted: 27 April 2025

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | *DOI: https://doi.org/10.48084/etasr.10738*

ABSTRACT

This study presents a Deep Learning (DL)-based approach for the early detection of Chronic Obstructive Pulmonary Disease (COPD) using a novel dual-branch Convolutional Neural Network (CNN) architecture.DL techniques are leveraged to recognize complex, early-stage patterns of the disease that may be overlooked by conventional medical assessments or traditional machine learning models, which are prone to misclassifying COPD as other lung conditions. To ensure robust model training, a pre-filtered dataset of lung sound recordings was used. These recordings, each 20 s in duration, were cleaned, standardized, and converted into two-dimensional representations using Mel spectrograms and Mel Frequency Cepstral Coefficients (MFCCs). These image-like features served as the input for the CNN model, enhancing its ability to distinguish COPD-specific acoustic patterns.To address the issue of class imbalance in the dataset, two data augmentation techniques, pitch-shifted noise injection and time-frequency masking, were applied, contributing to improved model generalization. The proposed CNN model achieved promising results, with a precision of 97.75%, an accuracy of 96.0%, a sensitivity of 97.96%, and an F1-score of 96.97% during validation. These performance metrics outperform those obtained from widely used CNN architectures, such as InceptionV3 and ResNet, highlighting the effectiveness of the proposed model. Overall, the proposed approach demonstrates significant potential as a reliable diagnostic support tool for early COPD detection.

Keywords-deep learning; chronic obstructive pulmonary disease; data augmentation; double-branch; convolutional neural network; mel frequency cepstral coefficients

I. INTRODUCTION

Respiratory diseases significantly impact health and are a major contributor to mortality. Among these, COPD is particularly concerning, ranking as the third leading cause of death globally, with an estimated three million fatalities in 2019 according to the World Health Organization (WHO) [1–3]. COPD is characterized by lung airflow obstruction due to structural changes in the alveoli,with common early symptoms including shortness of breath, chronic cough, and wheezing [4-6]. While COPD is preventable, it iscrucial to avoid or minimize exposure to risk factors, such as smoking and prolonged exposure to environmental pollutants, especiallyfine particulate matter under 2.5 microns in diameter($PM_{2.5}$) [7], and aerosols that have been shown to adversely affect lung health[8].

COPD typically manifests in two forms: emphysema, which damages the alveoli, and chronic bronchitis, which affects the airways [9]. The disease progressively diminishes respiratory function, severely impacting the quality of life and limiting physical activity. Socioeconomic factors also play a critical role, as individuals in underprivileged communities often face challenges in accessing timely diagnosis and treatment [10].

Therefore, an early diagnosis of COPD is of utmost importance to prevent irreversible lung damage. Currently, the standard diagnostic approach relies on auscultation by experienced physicians using a stethoscope, complemented by imaging techniques, such as chest X-rays, Magnetic Resonance Imaging (MRI), and Computed Tomography (CT) scans [11]. However, in many developing regions, limited access to healthcare professionals and the inherently subjective nature of auscultation can hinder timely diagnosis.

To address these limitations, this study proposes the development of a DL-based model for early COPD detection. The primary goal is to integrate the model into a micro-PC system paired with an electronic stethoscope, enabling an automated and accurate analysis of lung sounds. This setup could serve as a valuable diagnostic support tool, particularly in remote or resource-limited areas, and empower less experienced medical personnel to detect COPD at an early stage.

II. STATE OF THE ART AND WORKING MATERIAL FOR COPD DETECTION

A. State of the Art in COPD Detection

Recent advances in DL and machine learning have led to the development of various methods for identifying respiratory diseases, including COPD. These approaches typically rely on audio analysis of lung sounds and utilize sophisticated algorithms for feature extraction and classification. One notable method is described in [12], where MFCCs are used to generate spectrograms from lung sound recordings. These spectrograms are then analyzed using transfer learning techniques applied to models, such as MusicNN, OpenL3, and VGGish. The extracted features are further processed using Principal Component Analysis (PCA) and Support Vector Machines (SVM) to detect wheezing and crackles—key indicators of respiratory anomalies.Authors in [13] merged two lung sound databases and employed three entropy-based feature extraction techniques—Shannon entropy, logarithmic energy entropy, and spectral entropy. These features, derived from audio spectrograms, were used in conjunction with machine learning classifiers to categorize lung sounds into six classes, achieving promising classification accuracy. Authors in [14] implemented a hybrid model combining CNNs with Long Short-Term Memory (LSTM) networks to classify respiratory sounds into four categories. To address the challenge of class imbalance in their dataset, they applied the Focal Loss (FL) function, an improved version of the standard Cross-Entropy (CE) loss commonly used in image classification, enhancing the model's ability to learn from minority class examples.

B. Datasets

This study utilizes two prominent databases of lung sound recordings:

- ICBHI 2017 Dataset [15]: This publicly available dataset was created for the International Conference on Biomedical and Health Informatics (ICBHI) 2017 challenge. It comprises 920 audio recordings collected from 126 patients and includes labels for eight different respiratory conditions.
- KAUH Dataset [16]: The second dataset includes lung sound recordings from 112 patients at King Abdullah University Hospital (KAUH). These recordings are labeled into eleven respiratory condition categories. However, due to the limited sample size for each condition, this study focuses on a three-class classification task: distinguishing patients with COPD, healthy individuals, and those with other diseases.

Figure 1 presents a visual summary of the datasets used in this research, highlighting the number of samples and distribution across categories.

III. METHODOLOGY

The study development begins with the datasets used for the training and validation of the proposal algorithm, deploying the methodology shown in Figure 2. The proposed model was trained and validated using data augmentation techniques. All tests were conducted on a computer with an NVIDIA GeForce RTX 3060 GPU and 8 GB of RAM.

A. Preprocessing

In line with the objectives of this study, three primary categories were defined for classification:

- COPD: Representing patients diagnosed with COPD.
- Healthy: Denoting individuals with normal breathing patterns.
- Other Conditions: Encompassing respiratory diseases distinct from COPD.

To prepare the data for training, both datasets were merged, and the variability in audio durations was analyzed. A standardization process was necessary due to the presence of recordings with differing lengths. It was determined that a 20-s

duration would be optimal, as it reliably captures at least six full respiratory cycles, ensuring sufficient information for detecting respiratory anomalies. Recordings shorter than 15 s were excluded from the dataset. Only audios within the range of 15 to 25 s were retained, and those were either padded or trimmed to achieve the uniform 20-s length.

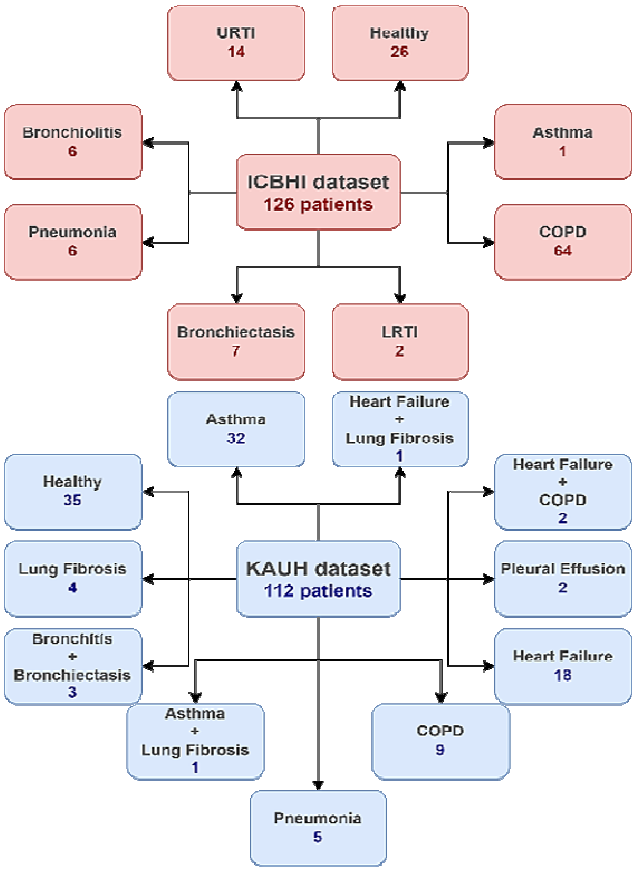

Fig. 1. Information about ICBHI and KAUH dataset.

Another critical aspect of preprocessing involved selecting the frequency range relevant to lung sound analysis. Lung sounds predominantly occupy the 50 Hz-2500 Hz bandwidth [17]. Therefore, a sixth-order Butterworth bandpass filter was applied to isolate this spectral range and enhance the Signal-to-Noise Ratio (SNR), particularly in the higher-frequency bands, where meaningful respiratory information resides. However, it is important to note that frequencies below 100 Hz often exhibit significant spectral overlap between heart and lung sounds [18]. As this overlap increases the complexity of isolating COPD-specific acoustic markers, this lower frequency band was excluded from analysis to improve the model's focus on relevant features for early-stage COPD detection.

B. Data Augmentation Criteria

To address the challenge posed by the class imbalance in the merged lung sound dataset, illustrated in Figure 3, two distinct data augmentation techniques were employed. Training and validation were performed independently for each augmentation method to allow a fair comparison of their effectiveness and to identify the most suitable approach for improving the model performance.

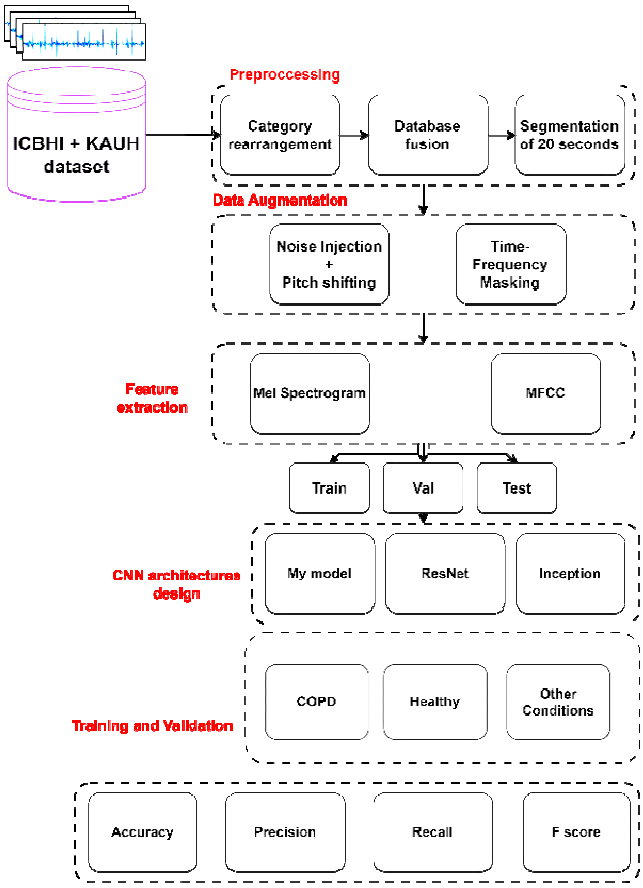

Fig. 2. Stages employed in this study.

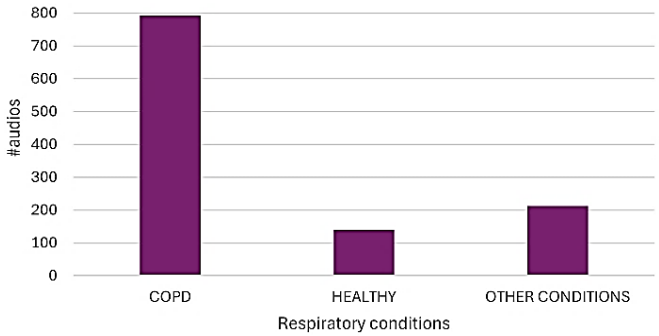

Fig. 3. Number of audio signals in each category after data fusion.

The augmentation strategies are summarized as follows:

- Tone shifting and the injection of two levels of Gaussian noise into the lung sound audio [19], with SNR levels of 40 and 30 dB.

- Time and frequency masking,a technique that involves randomly replacing a segment of the spectrogram band (Mel or MFCC) in the signal with silence [20].

As illustrated in Figure 4, the implementation of these augmentation techniques substantially increased the number of available training samples. The COPD and Other Conditions categories were both expanded to 500 samples each. However, despite these efforts, the Healthy category reached only 331 samples, indicating that a completely balanced dataset was not achieved. Nonetheless, this augmentation step significantly improved the dataset diversity and model generalizability.

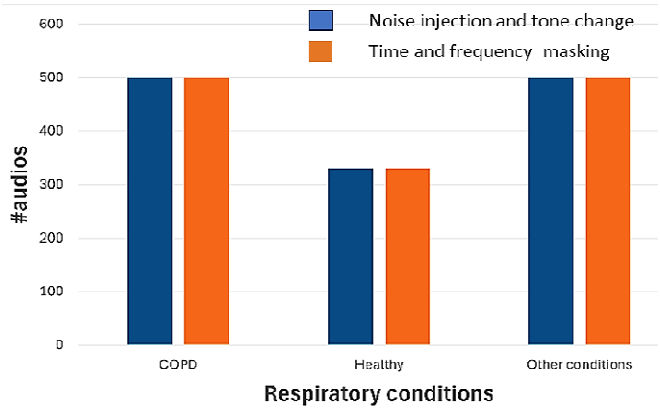

Fig. 4. Number of audio signals in each class after data augmentation for masking time and frequency, as well as noise injection and tone change.

C. Feature Extraction

Since CNNs are not designed to process raw audio signals directly, it is necessary to transform these signals into a suitable visual representation using a Mel spectrogram and the MFCC. These representations enable the use of CNNs by converting the audio data into image-like formats.

The MEL spectrogram uses the Discrete Fourier Transform (DFT), as shown in:

$$X[m] = \sum_{n=0}^{N-1} x[n] \cdot e^{-\frac{j2\pi nm}{N}} \quad (1)$$

where $x[n]$ denotes the n-th sample of the signal, whereas N signifies the total number of samples. In this regard, the DFT is an operation applicable to stationary signals. Still, because lung sounds are not stationary, it was necessary to apply the variant called Short-Time Fourier Transform (STFT), as expressed in:

$$X[m] = \sum_{n=0}^{L-1} x[n - mL] \cdot w[n] \cdot e^{-jwn} \quad (2)$$

STFT involves segmenting the signal into S samples of L size for which a window function $w[n]$ [21] is used. The Hamming, Hanning, and rectangular windows are the most commonly used.

Obtaining the Mel spectrogram requires using the Mel scale, as expressed in (3). A filter bank must also be used in the STFT block [22], as shown in Figure 5.

$$mel_f = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right) \quad (3)$$

Another feature extraction technique commonly deployed in audio is the MFCC, which requires a pre-emphasis stage, as well as an additional technique known as Discrete Cosine Transform (DCT). These stages are added at the Mel spectrogram stages' beginning and end, respectively, as shown in Figure 6. The pre-emphasis block acts as a high-pass filter, as expressed in (4). Additionally, the DCT is utilized to calculate the MFCC coefficients, described in (5) [23].

$$\hat{x}[n] = x[n] - \rho \cdot x[n-1] \quad (4)$$

$$c_n = \sum_{a=1}^{A} log(\hat{X}[a]) \cdot cos\left(n\left(a - \frac{1}{2}\right)\frac{\pi}{A}\right) \quad (5)$$

where $\hat{X}[a]$ represents the signal processed before the Mel-Filterbank stages [24].

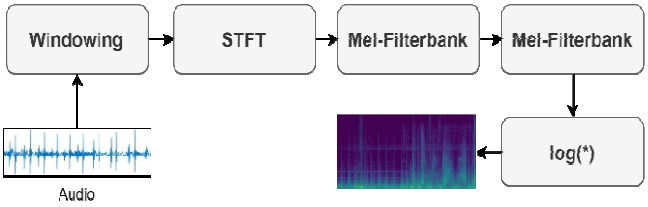

Fig. 5. Stages to obtain the Mel spectrogram.

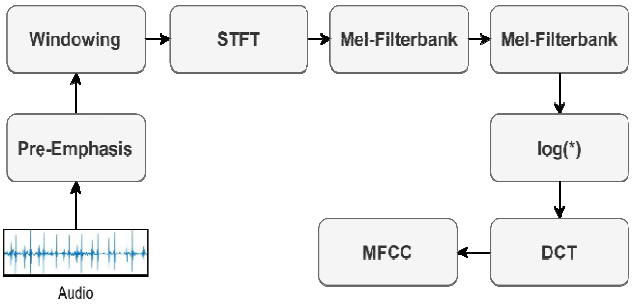

Fig. 6. Stages to obtain the MFCC representation.

D. Criteria for Designing CNN Architectures

CNN-based architectures utilize convolutional layers to automatically extract the most relevant features from input images. These extracted features form an input matrix for subsequent layers. The convolution operation itself is defined in:

$$(\Im * \mathcal{H})[n] = \sum_{l=-\infty}^{\infty} \Im[n-l]\mathcal{H}[l] \quad (6)$$

In this regard, the number of layers is not synonymous with greater precision. The higher the number of layers is, the more the model becomes over-parameterized and the more chaotic the operation is, with a strong tendency to overfit. In applications involving discontinuous and sparse visual features, such as those found in spectrogram representations of respiratory sounds, single-branch CNNs often fail to capture multiscale dependencies and relevant metrics. Therefore, it was necessary to work with a two-branch model capable of extracting features at different scales in parallel, each at the appropriate depth [25]. The two-branch CNN model used in this research corresponds to that of Figure 7.

Each network branch follows a block structure comprising three sets of operations. The first operation within each block consists of two convolutional layers. The first convolutional layer of a branch uses a 5x5 kernel, and the other uses a 3x3 kernel [26]. This implies that one branch aims to extract more meaningful features from the input, while the other extracts secondary features. After the convolutional layers, a Batch Normalization (BN) layer is inserted to standardize the inputs of each CNN layer. This promotes stable training dynamics and increases learning rates, resulting in faster training [27]. This algorithm uses a series of steps, as proposed in [28]. First, the mean and standard deviation of the input for each layer are calculated using:

$$\mu_B = \frac{1}{K}\sum_{j=1}^{K} x_j \quad (7)$$

$$\sigma_B^2 = \frac{1}{K}\sum_{j=1}^{K}\left(x_j - \mu_B\right)^2 \quad (8)$$

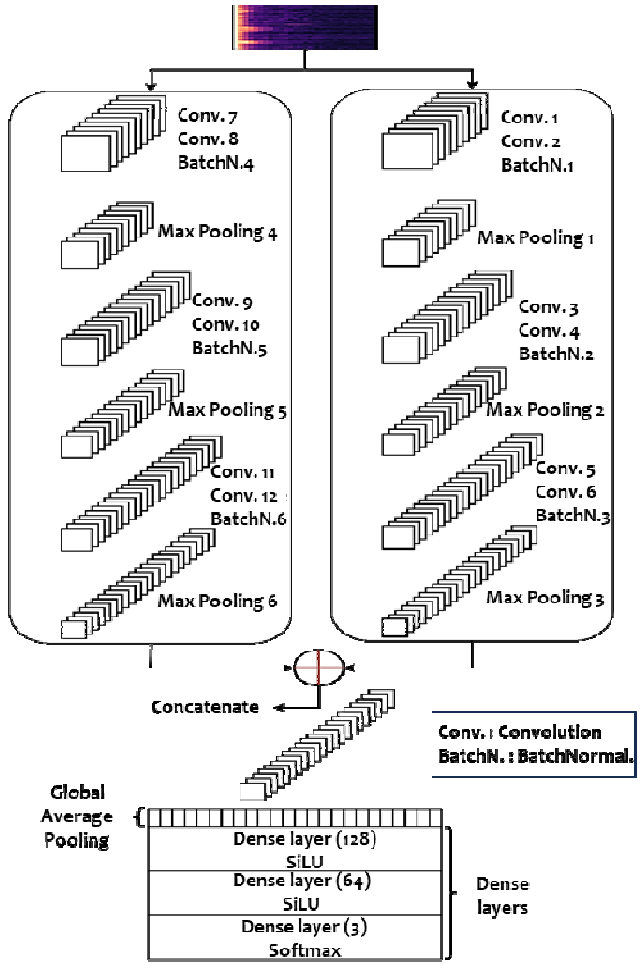

Fig. 7. Architecture of the proposed model for early detection of COPD using dual-branch CNN.

The inputs are normalized using (9); then, the normalized outputs are adjusted using (10), adding two parameters, $\delta$and φ, whose value must be obtained during the training process:

$$\widehat{x_J} = \frac{x_j - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \quad (9)$$

$$y_J = \delta\widehat{x_J} + \varphi \quad (10)$$

After the convolutional layers and BN, a max pooling layer is introduced to reduce the dimensionality of the feature maps.

Once both branches have extracted features, their outputs are concatenated, enabling the fusion of multiscale information into a unified representation. To convert the fused output into a one-dimensional vector for input to dense layers, Global Average Pooling (GAP) was employed. GAP outperformed the Flatten operation in both speed and model generalization by significantly reducing the number of learnable parameters.

In the final part of the model, dense layers of sizes 128, 64, and 3 are added, and the last layer has three units corresponding to the output classes. The Sigmoid Weighted Linear Unit (SiLU) activation function is used in the convolutional layers. The expression of the first two dense layers is shown in (11). Several activation functions were tested during the experimentation, including ReLU, SiLU, ReLU6, and Tanh from the Keras library. SiLU [29] exhibited superior performance among them due to its inherent harmonic reduction and lack of discontinuities, like ReLU and Tanh spectral density.

$$silu(x) = x \cdot \sigma(x) = x \cdot \frac{1}{1+e^{-x}} \quad (11)$$

For the output dense layer, the softmax activation function [30] classifies the three categories proposed in the present study: COPD, Healthy, and Other Conditions.

E. Pre-trained Architectures

1) Inception Net

The Inception architecture is a CNN developed by a team at Google. Its distinctive feature is the introduction of filter blocks with varying dimensions, organized into parallel branches that are merged to serve as input for the next layer [32]. This design achieved outstanding performance, surpassing the challenge posed by the ImageNet model. The architecture later evolved into variants, such as InceptionV3, which introduced enhancements to the original Inception module [33]. As is typical, the number of neurons increased with model capacity, resulting in approximately 23.9 million parameters.

The model enables multi-scale feature extraction through parallel convolutional layers with different kernel sizes, effectively capturing diverse patterns. Its structure supports deeper networks without suffering from vanishing gradients, leading to improved classification accuracy. Furthermore, its modular design enhances adaptability across tasks and helps reduce overfitting. Given its strong performance even with limited datasets, it has been considered a highly suitable candidate architecture.

2) ResNet20

The ResNet architecture is a CNN that introduced the concept of residual blocks, incorporating skip connections from previous layers. These connections ensure that the output features of a layer are at least as informative as those from

preceding layers [34]. The architecture has undergone continuous development, leading to variants, such as ResNet50 and ResNet100, which differ primarily in the number of convolutional layers and the corresponding increase in parameters.In the present study, the Keras library was used to implement various ResNet versions, with ResNet20 specifically tested for the detection of COPD. This architecture comprises approximately 5 million parameters.

The key benefit of residual connections is that if a particular layer does not enhance model performance, its contribution can be bypassed, preserving the integrity of the learning process. Furthermore, making these skip connections learnable allows their influence during training to be adaptively optimized. Network depth can also be tuned to balance model complexity and performance. Additionally, the application of regularization techniques, such as dropout or BN, can further improve generalization and model robustness.

F. Optimization Algorithms

To enhance the performance of the proposed model, four optimization algorithms were employed: Stochastic Gradient Descent (SGD), AdaDelta, Root Mean Square Propagation (RMSprop), and Adam, as described in [31].

1) Stochastic Gradient Descent

SGD is a variant of the batch gradient descent algorithm.It updates the weight vector$W$using:

$$W_{t+1} = W_t - \eta \nabla_W J(W) \tag{12}$$

In contrast to batch gradient descent, SGD performs an update for each training sample $x^{(i)}$,$y^{(i)}$:

$$W_{t+1} = W_t - \eta \nabla_W J\left(W, x^{(i)}, y^{(i)}\right) \tag{13}$$

where η is the learning rate and J is the cost function.

2) AdaDelta

AdaDelta adapts the learning rate using a moving window of accumulated gradients. The update equations are:

$$E[g^2]_t = \gamma E[g^2]_{t-1} + (1-\gamma) g_t^2 \tag{14}$$

$$\Delta W_t = -\frac{\sqrt{E[\Delta W^2]_t + \varepsilon}}{\sqrt{E[g^2]_t + \varepsilon}} \odot g_t \tag{15}$$

$$W_{t+1} = W_t + \Delta W_t \tag{16}$$

where $E[g^2]_t$ is the exponentially decaying average of past squared gradients,andγ is a hyperparameter set to 0.9.

3) RMSprop

RMSprop is designed for efficient optimization on large datasets. It modifies AdaDelta by using:

$$[g^2]_t = \beta E[g^2]_{t-1} + (1-\beta) \nabla_W J(W_t)^2 \tag{17}$$

$$W_t = W_{t-1} - \frac{\eta}{\sqrt{E[g^2]_t}} * \nabla_W J(W_t) \tag{18}$$

where$\nabla_W J(W_t)$ is the gradient of the cost function, and β is the moving average parameter.

4) Adam

Adam combines the advantages of both RMSprop and momentum. It computes adaptive learning rates for each parameter:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) * \nabla_W f_t(W_{t-1}) \tag{19}$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) * \nabla_W f_t(W_{t-1})^2 \tag{20}$$

Bias-corrected estimates are then calculated as:

$$\tilde{m}_t = m_t (1-\beta_1^t)^{-1} \tag{21}$$

$$\tilde{v}_t = v_t (1-\beta_2^t)^{-1} \tag{22}$$

The final weight update is performed using:

$$W_t = W_{t-1} - \alpha \tilde{m}_t / (\sqrt{\tilde{v}_t} + \lambda) \tag{23}$$

where by default $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.99$, and $\lambda = 10^{-8}$.

IV. RESULTS OF THE DL MODELS TESTED FOR THE DETECTION OF COPD

A. Results of the Double-Branch Model Test

The results are presented in two sets, based on the data augmentation techniques applied. In each case, the models were evaluated using both MEL spectrograms andMFCCs. The objective was to identify the model and determine the conditions under which COPD is best detected.

Figure 8 illustrates the performance of the dual-branch CNN model using a dataset augmented with tone shifting and noise injection. The analysis of the performance curves indicates that the model achieved high accuracy when combined with the proposed optimization strategies and feature extraction methods. Among the configurations tested, the highest validation accuracy was obtained using MFCCs with the SGD and RMSprop optimizers, achieving 97.9% and 97.5% accuracy, respectively.

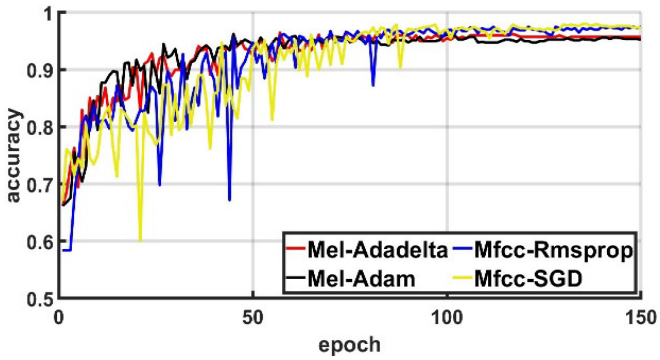

Fig. 8. Validation accuracy results for the double-branch CNN model, using tone change and noise injection for data augmentation.

Figure 9 presents the model's performance using time and frequency masking for data augmentation. In this scenario, accuracy was generally lower than in the previous case. The best result was achieved using MEL spectrogram features in

combination with the Adam optimizer, reaching a validation accuracy of 92.2%.

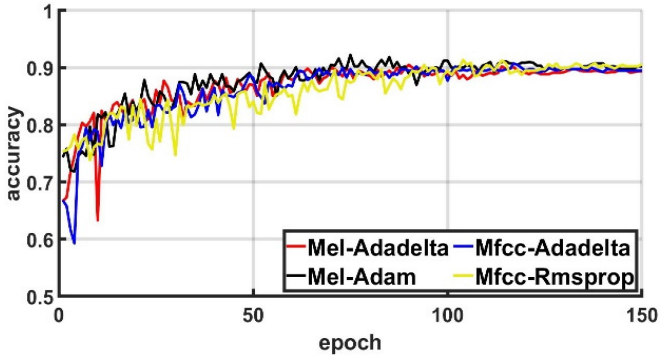

Fig. 9. Validation accuracy results for the dual-branch CNN model using the time and frequency masking criterion for data augmentation.

B. Results of the Pre-Trained Models Tested to Detect COPD

Figure 10 presents the results of the COPD detection using the InceptionV3 network. This CNN was trained on data augmented through noise injection and pitch shifting, with MFCCs used for feature extraction. The model was evaluated utilizing various optimization algorithms, with Adam and RMSprop yielding the best performance, achieving validation accuracies of 96.6% and 96.5%, respectively.

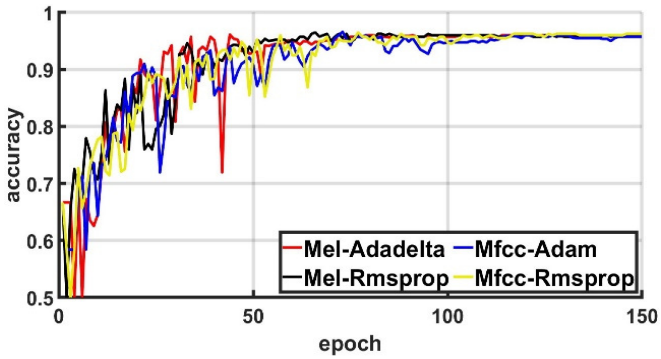

Fig. 10. Validation accuracy results from the InceptionV3 model using the criterion change tone and noise injection for data augmentation.

Figure 11 displays the accuracy results obtained with the InceptionV3 network using data augmented through time and frequency masking. In this configuration, the highest accuracy of 93.3% was achieved using MFCC features combined with the SGD optimizer.

Figure 12 illustrates the performance of the ResNet20 network trained on data augmented with pitch shifting and noise injection. The best result was obtained using MFCC features and the RMSprop optimizer, reaching an accuracy of 97.5%.

Figure 13 shows a decline in ResNet20's performance when time and frequency masking were used for data augmentation. In this case, the highest accuracy of 90.2% was achieved using MEL spectrogram features and the Adam optimizer.

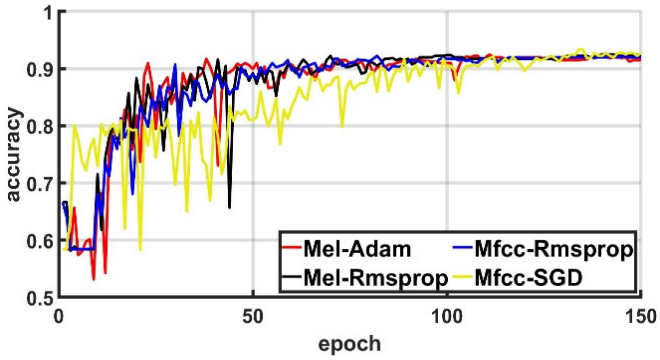

Fig. 11. Validation accuracy results from the InceptionV3 model using the criterion of time and frequency masking for augmentation data.

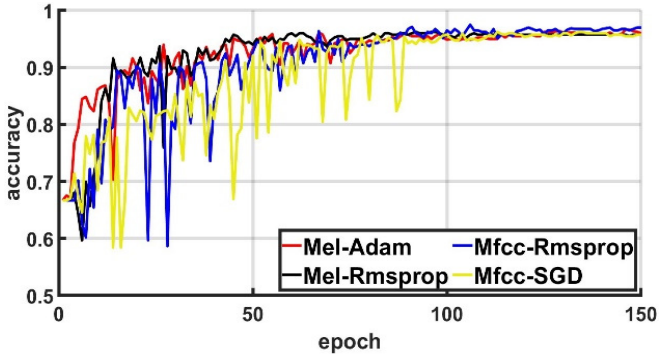

Fig. 12. Validation accuracy results from the ResNet20 model applying tone change and noise injection for data augmentation.

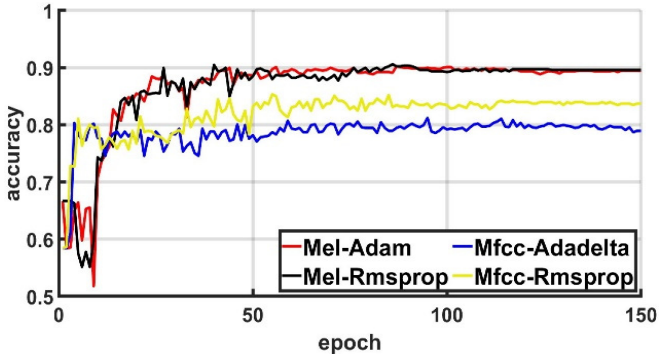

Fig. 13. Validation accuracy results from the ResNet20 model, using time and frequency masking for data augmentation.

V. ANALYSIS AND DISCUSSION OF THE RESULTS

Overall, the models trained with data augmentation based on time-frequency masking outperformed those using pitch shifting. However, accuracy alone does not provide a comprehensive assessment of the model performance.

Given the class imbalance in the dataset, it is more informative to consider the sensitivity, precision, and F1-score metrics, calculated by:

$$Acc. = \frac{TP+TN}{TP+FP+TN+FN} \quad (24)$$

$$Sensit. = \frac{TP}{TP+FN} \quad (25)$$

$$Precis. = \frac{TP}{TP+FP} \quad (26)$$

$$F1 - score = 2 \cdot \frac{Precis. \cdot Sensit.}{Precis. + Sensit.} \quad (27)$$

Table I summarizes the evaluation results, including accuracy, sensitivity, precision, and F1-score, allowing for a meaningful comparison across models.

TABLE I. COMPARATIVE METRIC RESULTS ACROSSTHE MODELS USED IN THIS STUDY

Model	Performance evaluation (%)			
	Acc.	Sensit.	Precis.	F1-score
Double-branch CNN + Mel + Augmentation 1 + Adadelta	96.5	94.7	94.7	94.5
Double-branch CNN + MFCC + Augmentation 1 + SGD	**97.75**	**96.0**	**97.96**	**96.97**
Double-branch CNN + Mel + Augmentation 2 + Adam	92.2	88.3	88.4	87.0
Double-branch CNN + Mel + Augmentation 2 + Adam	92.0	87.9	88.0	87.1
InceptionV3 + Mel + Augmentation 1 + Rmsprop	96.5	94.7	94.7	94.7
InceptionV3 + MFCC + Augmentation 1 + Adam	96.6	96.6	95.1	94.9
InceptionV3 + Mel + Augmentation 2 + Adam	96.5	94.7	94.7	94.5
InceptionV3 + MFCC + Augmentation 2 + SGD	92.5	88.7	88.7	88.4
ResNet20 + Mel + Augmentation 1 + Adam	96.6	94.7	95.1	94.9
ResNet20 + MFCC + Augmentation 1 + Rmsprop	97.5	96.2	96.2	96.4

The analysis of these metrics indicates that the proposed dual-branch CNN model using MFCC features outperforms the pre-trained architectures. This is especially evident when trained with pitch-shifted data and optimized using SGD, achieving 97.75% accuracy, 96.0% sensitivity, 97.96% precision, and a 96.97% F1-score.The proposed model achieves superior performance using significantly fewer parameters (297,443) compared to deeper networks, such as ResNet20 (5 million parameters) and InceptionV3 (22 million parameters). These results highlight the efficiency of the dual-branch architecture in spectrogram-based classification tasks, delivering high accuracy with reduced computational cost. A comparative analysis between the performance of the proposed double-branch models against models from other studies is shown in Table II.While the proposed models rank among the top-performing methods in terms of overall evaluation metrics, they slightly trail the fine Gaussian SVM and the boosted decision tree model in accuracy. However, they outperform the latter in terms of F1-score, indicating a better balance between precision and recall.

TABLE II. RESULTS FROM THE DUAL BRANCH MODELS COMPARED WITH OTHER MODELS FROM SIMILAR STUDIES

Model	Categories	Performance evaluation (%)		
		Acc.	Sensit.	F1-Score
Double-branch CNN + Mel + Augmentation 1 + Adadelta	**COPD, Healthy, and Other Conditions**	**96.5**	**94.7**	**94.5**
Double-branch CNN + MFCC + Augmentation 1 + SGD	**COPD, Healthy, and Other Conditions**	**97.75**	**96.0**	**96.97**
VGGish [2]	Wheezes and Crackles	81	---	---
Boosted decision tree [3]	Normal, Asthma, Heart Failure, Pneumonia, BRON, COPD	98.27	95.28	93.61
CNN-LSTM + FL [4]	Normal, Crackles, Wheezes, Crackles + Wheezes	76.39	52.78	68.52
Fine Gaussian SVM [5]	Asthma, Bronchiectasis, Bronchiolitis, COPD, Healthy, LRTI, Pneumonia, Upper Respiratory Infection	99	99.04	---

Table III presents the classification results (observation matrix) obtained from the MFCC-based dual-branch model using the SGD optimizer, while Figure 14 illustrates the corresponding confusion matrix. The model demonstrated strong classification performance, correctly identifying the majority of cases, with only 4 COPD, 7 Healthy, and 11 Other condition instances misclassified out of a total of 267 samples.

TABLE III. OBSERVATION MATRIX

Classes	Measures			
	TP	TN	FP	FN
COPD	96	165	2	4
Healthy	65	195	5	2
Other conditions	94	162	5	6

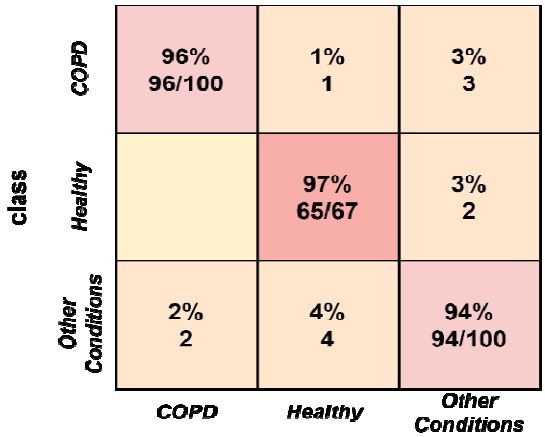

Fig. 14. Confusion matrix.

VI. CONCLUSION

This study proposed a novel dual-branch Convolutional Neural Network (CNN) architecture for the early detection of Chronic Obstructive Pulmonary Disease (COPD) using lung sound recordings. By leveraging Deep Learning (DL) and two-dimensional audio representations, Mel Frequency Cepstral coefficients (MFCCs) and Mel spectrograms, the model aimed to identify subtle acoustic patterns associated with COPD that might be missed by single-branch networks.

To ensure a robust and generalizable performance, the model incorporated data augmentation techniques, pitch-shifted noise injection, and time-frequency maskingto address class imbalance and improve training quality. The dual-branch design was introduced to mitigate the saw-tooth behavior observed in the training dynamics of the single-branch CNN. This architecture allowed the model to evaluate performance across two parallel data streams and dynamically select the one yielding better results, enhancing both stability and accuracy during validation.The final model, optimized using the Stochastic Gradient Descent (SGD) algorithm and evaluated on MFCC inputs, achieved an impressive accuracy of 97.75%, with a sensitivity of 96.0%, a precision of 97.96%, and an F1-score of 96.97%. These results demonstrate a clear improvement over conventional CNN models, such as ResNet and InceptionV3.

Beyond the performance metrics, the dual-branch CNN holds a strong practical value as a diagnostic support tool, particularly in underserved or rural regions, like those in Peru and other parts of Latin America, where access to pulmonologists is limited. The integration of this model into a portable system with an electronic stethoscope could empower non-specialist healthcare providers to detect COPD at an early stage, contributing to timely intervention and improved patient outcomes.

ACKNOWLEDGMENT

The authors thank the Dirección General de Investigación (DIGI-UNI) for support in this study, and the Facultad de IngenieríaEléctrica y Electrónica (FIEE-UNI) of the Universidad Nacional de Ingeniería for providing the necessary hardware and their kind attention to using their facilities, allowing the development and testing of the model proposed in this research.

REFERENCES

[1] World Health Organization, *World Health Statistics 2021: Monitoring Health for the SDGs, Sustainable Development Goals*, 1st ed. Geneva: World Health Organization, 2021.

[2] T. Siddiqui, M. Latif, M. U. Farooq, M. A. Baig, and Y. S. Hassan, "Chronic Obstructive Pulmonary Disease Diagnosis with Bagging Ensemble Learning and ANN Classifiers,"*Engineering, Technology & Applied Science Research*, vol. 14, no. 3, pp. 14741–14746, Jun. 2024, https://doi.org/10.48084/etasr.7106.

[3] R. Karla and R. Yalavarthi, "A Hybrid RNN-based Deep Learning Model for Lung Cancer and COPD Detection,"*Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 16847–16853, Oct. 2024, https://doi.org/10.48084/etasr.8181.

[4] J. Y. Choi and C. K. Rhee, "Diagnosis and Treatment of Early Chronic Obstructive Lung Disease (COPD),"*Journal of Clinical Medicine*, vol. 9, no. 11, Oct. 2020, Art. no. 3426, https://doi.org/10.3390/jcm9113426.

[5] P. J. Patel *et al.*, "Performance analysis of deep learning algorithms for classifying chronic obstructive pulmonary disease,"*Journal of Integrated Science and Technology*, vol. 12, no. 2, 2024, Art. no. 745.

[6] ADVOCATEHEALTH. "Chronic obstructive pulmonary disease (COPD)." Advocate Health Care. [Online]. Available: https://www.advocatehealth.com/health-services/lung-respiratory-care/chronic-obstructive-pulmonary-disease-copd.

[7] M. Elbarbary *et al.*, "Ambient air pollution, lung function and COPD: cross-sectional analysis from the WHO Study of AGEing and adult health wave 1,"*BMJ Open Respiratory Research*, vol. 7, no. 1, Dec. 2020, Art. no. e000684, https://doi.org/10.1136/bmjresp-2020-000684.

[8] O. R. Sánchez-Ccoyllo, C. G. Ordoñez-Aquino, J. Arratea-Morán, N. S. Marín-Huachaca, and W. Reátegui-Romero, "Describing Aerosol and Assessing Health Effects in Lima, Peru,"*International Journal of Environmental Science and Development*, vol. 12, no. 12, pp. 355–362, 2021, https://doi.org/10.18178/ijesd.2021.12.12.1361.

[9] V. Kim and G. J. Criner, "Chronic Bronchitis and Chronic Obstructive Pulmonary Disease,"*American Journal of Respiratory and Critical Care Medicine*, vol. 187, no. 3, pp. 228–237, Feb. 2013, https://doi.org/10.1164/rccm.201210-1843CI.

[10] Y. S. Lee, J. Y. Oh, K. H. Min, S. Y. Lee, K. H. Kang, and J. J. Shim, "The association between living below the relative poverty line and the prevalence of chronic obstructive pulmonary disease,"*Journal of Thoracic Disease*, vol. 11, no. 2, pp. 427–437, Feb. 2019, https://doi.org/10.21037/jtd.2019.01.40.

[11] S. W. Ali, M. Asif, M. Rashid, S. Tanvir, S. Shams, and S. Abid, "Detection of Crackle and Wheeze in Lung Sound using Machine Learning Technique for Clinical Decision Support System,"*VAWKUM Transactions on Computer Sciences*, vol. 11, no. 1, pp. 67–78, Mar. 2023, https://doi.org/10.21015/vtcs.v11i1.1384.

[12] S. Ali, S. Tanweer, S. Khalid, and N. Rao, "Mel Frequency Cepstral Coefficient: A Review," in *Proceedings of the 2nd International Conference on ICT for Digital, Smart, and Sustainable Development, ICIDSSD 2020, 27-28 February 2020, Jamia Hamdard, New Delhi, India*, New Delhi, India, 2021, https://doi.org/10.4108/eai.27-2-2020.2303173.

[13] L. Fraiwan, O. Hassanin, M. Fraiwan, B. Khassawneh, A. M. Ibnian, and M. Alkhodari, "Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers,"*Biocybernetics and Biomedical Engineering*, vol. 41, no. 1, pp. 1–14, Jan. 2021, https://doi.org/10.1016/j.bbe.2020.11.003.

[14] G. Petmezas *et al.*, "Automated Lung Sound Classification Using a Hybrid CNN-LSTM Network and Focal Loss Function,"*Sensors*, vol. 22, no. 3, Feb. 2022, Art. no. 1232, https://doi.org/10.3390/s22031232.

[15] Z. Sun, "ICBHI 2017 challenge." Harvard Dataverse, 2023, Art. no 1978998275, https://doi.org/10.7910/DVN/HT6PKI.

[16] M. Fraiwan, L. Fraiwan, B. Khassawneh, and A. Ibnian, "A dataset of lung sounds recorded from the chest wall using an electronic stethoscope,"*Data in Brief*, vol. 35, Apr. 2021, Art. no. 106913, https://doi.org/10.1016/j.dib.2021.106913.

[17] S. Reichert, R. Gass, C. Brandt, and E. Andrès, "Analysis of Respiratory Sounds: State of the Art,"*Clinical medicine. Circulatory, respiratory and pulmonary medicine*, vol. 2, Jan. 2008, Art. no. CCRPM.S530, https://doi.org/10.4137/CCRPM.S530.

[18] T. H. Falk, W.Y. Chan, E. Sejdic, and T. Chau, "Spectro-Temporal Analysis of Auscultatory Sounds," in *New Developments in Biomedical Engineering*, D. Campolo, Ed. InTech, 2010.

[19] M. E. Akbiyik, "Data Augmentation in Training CNNs: Injecting Noise to Images." arXiv, 2023, https://doi.org/10.48550/ARXIV.2307.06855.

[20] G. Zhou, Y. Chen, and C. Chien, "On the analysis of data augmentation methods for spectral imaged based heart sound classification using convolutional neural networks,"*BMC Medical Informatics and Decision Making*, vol. 22, no. 1, Aug. 2022, Art. no. 226, https://doi.org/10.1186/s12911-022-01942-2.

[21] E. Flórez, S. Cardona, and L. Jordi, "Selecting the temporal window in the short time Fourier transforms used in the vibration signal analysis to determine flats at train's wheels,"*Revista Facultad de Ingeniería*

Universidad de Antioquia, no. 50, pp. 145–158, Mar. 2013, https://doi.org/10.17533/udea.redin.14940.

[22] S. Guha, A. Das, P. K. Singh, A. Ahmadian, N. Senu, and R. Sarkar, "Hybrid Feature Selection Method Based on Harmony Search and Naked Mole-Rat Algorithms for Spoken Language Identification From Audio Signals,"*IEEE Access*, vol. 8, pp. 182868–182887, 2020, https://doi.org/10.1109/ACCESS.2020.3028121.

[23] B. Barai, D. Das, N. Das, S. Basu, and M. Nasipuri, "VQ/GMM-Based Speaker Identification with Emphasis on Language Dependency," in *Advanced Computing and Systems for Security*, vol. 883, R. Chaki, A. Cortesi, K. Saeed, and N. Chaki, Eds. Singapore: Springer Singapore, 2019, pp. 125–141.

[24] R. Mushi and Y.-P. Huang, "Assessment of Mel-Filter Bank Features on Sound Classifications Using Deep Convolutional Neural Network," in *2021 International Conference on System Science and Engineering (ICSSE)*, Ho Chi Minh City, Vietnam, Aug. 2021, pp. 334–339, https://doi.org/10.1109/ICSSE52999.2021.9538433.

[25] M. Del Coco, P. Carcagnì, M. Leo, P. Spagnolo, P. L. Mazzeo, and C. Distante, "Multi-branch CNN for Multi-scale Age Estimation," in *Image Analysis and Processing - ICIAP 2017*, vol. 10485, S. Battiato, G. Gallo, R. Schettini, and F. Stanco, Eds. Cham: Springer International Publishing, 2017, pp. 234–244.

[26] A. Ganjdanesh, S. Gao, and H. Huang, "EffConv: Efficient Learning of Kernel Sizes for Convolution Layers of CNNs," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Washington, D.C., USA, Jun. 2023, vol. 37, pp. 7604–7612, https://doi.org/10.1609/aaai.v37i6.25923.

[27] C. Garbin, X. Zhu, and O. Marques, "Dropout vs. batch normalization: an empirical study of their impact to deep learning,"*Multimedia Tools and Applications*, vol. 79, no. 19–20, pp. 12777–12815, May 2020, https://doi.org/10.1007/s11042-019-08453-9.

[28] W. M. Fatihia, A. Fariza, and T. Karlita, "CNN with Batch Normalization Adjustment for Offline Hand-written Signature Genuine Verification,"*JOIV : International Journal on Informatics Visualization*, vol. 7, no. 1, Feb. 2023, Art. no. 200,https://doi.org/10.30630/joiv.7.1.1443.

[29] M. Islam, S. S. Arora, R. Chatterjee, P. Rindal, and M. Shirvanian, "Compact: Approximating Complex Activation Functions for Secure Computation,"*Proceedings on Privacy Enhancing Technologies*, vol. 2024, no. 3, pp. 25–41, Jul. 2024, https://doi.org/10.56553/popets-2024-0065.

[30] S. Mehra, G. Raut, R. D. Purkayastha, S. K. Vishvakarma, and A. Biasizzo, "An Empirical Evaluation of Enhanced Performance Softmax Function in Deep Learning,"*IEEE Access*, vol. 11, pp. 34912–34924, 2023, https://doi.org/10.1109/ACCESS.2023.3265327.

[31] C. Peel and T. K. Moon, "Algorithms for Optimization [Bookshelf],"*IEEE Control Syst.*, vol. 40, no. 2, pp. 92–94, Apr. 2020, https://doi.org/10.1109/MCS.2019.2961589.

[32] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9, https://doi.org/10.1109/CVPR.2015.7298594.

[33] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2818–2826, https://doi.org/10.1109/CVPR.2016.308.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, https://doi.org/10.1109/CVPR.2016.90.