Journal of Robotics and Control (JRC)
Volume 5, Issue 2, 2024
ISSN: 2715-5072, DOI: 10.18196/jrc.v5i2.21422

Enhancing Pulmonary Disease Classification in Diseases: A Comparative Study of CNN and Optimized MobileNet Architectures

Omar Nadhim Mohammed [1*]
[1] School of Information Technologies, Altınbaş Üniversitesi, İstanbul, Turkey
Email: [1] omar.n1dhem@gmail.com
*Corresponding Author

***Abstract*—Background: Deep learning technologies, especially Convolutional Neural Networks (CNNs), are revolutionizing the field of medical imaging by providing advanced tools for the accurate classification of pulmonary diseases from chest X-ray (CXR) images. In our study, we employed both traditional CNN models and MobileNet architectures to classify various chest diseases using CXR images. Initially, a conventional CNN model was utilized to estab- lish a baseline accuracy. Subsequently, we adopted MobileNet, known for its efficiency in processing image data, to enhance classification performance. To further optimize the system, we applied Energy Valley Optimization (EVO) for hyperparameter tuning. The baseline CNN model achieved an accuracy of 85.91%. The implementation of MobileNet significantly improved this metric, reaching a pre-optimization accuracy of 93.30%. Post-EVO optimization, the accuracy was further enhanced to 94.18%. Comparative analysis of accuracy, precision, recall, F1-score, and ROC curves was conducted to illustrate the impact of hyperparameter tuning on model performance in medical diagnostics. Our findings demonstrate that while standard CNNs provide a solid foundation for CXR image classification, the integration of MobileNet architectures and EVO for hyperparameter adjustment significantly boosts diagnostic accuracy. This advancement in automated medical image analysis could potentially transform the landscape of pulmonary disease diagnosis, offering a more robust framework for accurate and efficient patient care.**

Keywords—Deep Learning; Convolutional Neural Networks; MobileNet; Chest X-Ray Classification; Hyperparameter Optimization; Energy Valley Optimization; Medical Imaging; ROC Curve.

I. INTRODUCTION

As 2019 drew to a close, the world began grappling with the outbreak of a new coronavirus disease, subsequently named COVID-19. This virus, which emerged in the city of Wuhan in Eastern China in December 2019, quickly escalated into a global crisis. By early 2020, the World Health Organization (WHO) declared it a "Public health emergency of international concern," and by March 2020, it was recognized as a pandemic [1]. By March 2021, COVID-19 had affected approximately 118.7 million people globally, resulting in around 2.6 million deaths. The virus primarily manifests as pneumonia, accompanied by symptoms like fatigue, dry cough, and fever. One of the main diagnostic tools is the reverse transcription polymerase chain reaction (RT-PCR), which involves testing respiratory samples. The results from this test can take from a few hours up to two days to process. Despite its reliability, this method is both costly and time-consuming [2]. Consequently, researchers are actively seeking alternative methods for virus detection. As of the current knowledge, there is no definitive medical treatment specifically for COVID-19 [3].

The advancement of artificial intelligence (AI) has significantly revolutionized the diagnosis of various diseases, with its remarkable capabilities in automating image classification tasks. This is largely attributed to the various machine learning techniques, which are designed to learn and make informed decisions from extensive datasets. AI operates by analyzing this data, leading to its application in tasks traditionally requiring human intelligence, such as speech recognition, language translation, and visual perception [4], [5].

Deep learning, a subset of machine learning, is particularly noteworthy for its focus on automatic feature extraction and image classification. This approach has been instrumental in numerous fields, especially in healthcare. It enables the creation of models that offer precise predictions and classifications of different diseases based solely on image analysis. These diseases include, but are not limited to, breast cancer [6], liver ailments [7], colon cancer [8], brain tumors [9], skin cancer [10], lung cancer [11], pneumonia [12], and more recently, in the diagnosis of COVID-19. Deep learning distinguishes itself from traditional machine learning by its ability to form more abstract representations of data as the network's depth increases. This capability allows for automatic feature extraction, leading to models that are not only more accurate but also require minimal to no human intervention. Unlike conventional machine learning algorithms that need specified features, deep learning algorithms derive these features through a sequence of non-linear functions, intricately combined to optimize model accuracy.

Recent research in the field of deep learning has focused extensively on its application in classifying COVID-19 using various imaging techniques. A significant body of work has explored the use of chest X-rays for this purpose, as highlighted in studies [13]–[15]. Additionally, there's a growing interest in employing computed tomography (CT) scans for similar diagnostic purposes, with several studies [16]–[22] delving into this area.

Another avenue of research has been the detection and diagnosis of COVID-19 using lung dataset analyses, as shown in studies [5], [23]. Some of these studies have implemented convolutional neural networks (CNNs), albeit with limited datasets, to classify and detect COVID-19 from chest X-ray images [24], [25]. There's also a notable focus on distinguishing COVID-19 from other chest-related diseases like pneumonia, with several studies [26]–[30] dedicated to this differentiation. Importantly, research indicated in Ref. [31] suggests that while chest X-rays may be less effective in the initial stages of COVID-19, CT scans can be useful even before symptoms manifest [32]. A challenge in diagnosing using chest CT or X-ray images lies in the potential overlap of symptoms between COVID-19, pneumonia, and chest cancer, especially when the diagnosis is made by less experienced individuals or in the absence of a comprehensive patient history. This underscores the need for automating the diagnostic process, ensuring accurate identification of these diseases.

The utilization of deep learning in medical imaging is primarily motivated by the urgent need to enhance diagnostic accuracy in the detection of pulmonary diseases [33]–[40], a need that has become even more pressing in the wake of the COVID-19 pandemic. Traditional diagnostic methods, while effective, often require significant time and resources, leading to delays in treatment and increased risk of disease transmission. In contrast, deep learning offers a faster, more efficient alternative by leveraging complex algorithms to analyze and interpret medical images with a high degree of precision [41]–[54]. This technology not only promises to expedite the diagnostic process but also aims to improve the accuracy of diagnoses, thereby facilitating timely and appropriate medical interventions. The potential of deep learning to automate and refine the analysis of medical images represents a significant advancement in the fight against pulmonary diseases, including COVID-19, by providing healthcare professionals with powerful tools to detect and differentiate between various conditions quickly and accurately. The advent of this technology marks a pivotal moment in medical diagnostics, where the emphasis on speed and precision can significantly impact patient outcomes and public health at large.

The COVID-19 pandemic, first identified in Wuhan, China in December 2019, rapidly escalated from a regional outbreak to a global crisis, with the World Health Organization declaring it a Public Health Emergency of International Concern by January 2020 and a pandemic by March 2020. As of March 2021, the virus had affected nearly 118.7 million people and claimed approximately 2.6 million lives worldwide [55], manifesting not only as a health crisis but also as a significant disruptor to economies and societies at large. The urgency for efficient diagnostic methods became paramount as the standard RT-PCR testing, despite its reliability, posed limitations due to its time-consuming and costly nature, which led to diagnostic bottlenecks and strained healthcare systems globally.

In this climate, the potential of Artificial Intelligence (AI) and deep learning became evident, offering a beacon of hope for addressing these diagnostic challenges. AI's capacity to quickly analyze vast datasets and extract complex patterns promised to accelerate and enhance the precision of COVID-19 diagnostics. Beyond image classification, AI's broader impact on healthcare is multifaceted, extending to diagnostics, treatment planning, and patient care, revolutionizing traditional approaches with its predictive analytics and automation capabilities. Specifically, deep learning's prowess in medical imaging is underscored by its automated feature extraction, which has been pivotal in improving disease classification accuracy, as evidenced by various successful applications documented in recent literature. Our research nestles within these advancements, aiming to bridge a critical gap by enhancing the accuracy and efficiency of COVID-19 detection from chest X-ray images. This study not only endeavors to bolster the fight against the current pandemic through improved diagnostic tools but also contributes to the foundation of AI in healthcare, setting the stage for future innovations that could reshape patient outcomes and public health strategies in the wake of such unprecedented global challenges.

II. RELATED WORK

In the field of medical imaging, particularly for COVID-19 detection, deep learning has shown significant promise, with several studies utilizing various methodologies and datasets to achieve high accuracy levels.

A notable study [56] applied three different convolutional neural network (CNN) models for detecting COVID-19 pneu- monia in patients using chest X-ray radiographs. These models, particularly ResNet50, demonstrated high classification accuracy (98%) without the need for feature extraction or selection. Similarly, Ref. [27] employed a generative adversarial network (GAN) coupled with fine-tuned deep transfer learning approaches like AlexNet, GoogLeNet, Squeeznet, and Resnet18, specifically for chest X-ray image analysis. The use of GANs helped mitigate overfitting issues and expanded the dataset of 5863 X-ray images.

In another approach, Ref. [57] used digital X-ray images to differentiate between COVID-19 and pneumonia patients. This study utilized four pre-trained CNNs: ResNet18, AlexNet, SqueezeNet, and DenseNet201, and employed image augmentation techniques to enlarge the dataset for training. The accuracy achieved was 98 CT scans have also been instrumental in COVID-19 detection. Ref. [58] used transverse section CT images to distinguish COVID-19 patients from others, such as those with influenza-A. This study involved 618 CT samples and utilized two three-dimensional (3D) CNN models, achieving an accuracy of 86.7%.

Ref. [16] introduced a weakly-supervised deep learning-based model named DeCoVNet, which was trained using segmented lung regions from CT images. This model reached a detection accuracy of 90.1%. Additionally, a CT scan COVID- 19 dataset was compiled in Ref. [17] to aid in virus detection. The authors used this dataset to train a deep CNN model, achieving an accuracy of 84.7%.

Ref. [28] proposed a rapid COVID-19 diagnosis method using CNNs applied to CT images, where the highest accuracy (99.51%) was achieved by ResNet-101.

In the realm of chest X-ray analysis, Ref. [59] proposed a classification schema that included both multiclass and hierarchical classifications for identifying COVID-19 and pneumonia cases, achieving an F1-Score of 65% and 0.89, respectively.

Lastly, Ref. [60] focused on detecting COVID-19 from chest X-rays using a small dataset. The study utilized pre-trained ResNet50 and VGG-16 models alongside their CNN, yielding an accuracy of 91.24%.

These varied approaches underline the versatility and potential of deep learning in medical imaging, particularly in the timely and accurate detection of COVID-19 using both X-ray and CT images.

TABLE I. SUMMARY OF RELATED WORK ON DEEP LEARNING FOR COVID-19 DETECTION

Reference	Method	Image Type	Accuracy
Ref. [56]	ResNet50, InceptionV3, Inception-ResNetV2	Chest X-ray	98%, 97%, 87%
Ref. [27]	GAN with AlexNet, GoogLeNet, Squeeznet, Resnet18	Chest X-ray	N/A
Ref. [57]	ResNet18, AlexNet, SqueezeNet, DenseNet201	Chest X-ray	98%
Ref. [58]	3D CNN models (ResNet23, etc.)	CT	86.7%
Ref. [16]	DeCoVNet (Weakly- supervised model)	CT	90.1%
Ref. [17]	Deep CNN model (Transfer learning)	CT	84.7%
Ref. [28]	Various CNNs including ResNet-101	CT	99.51%
Ref. [59]	Multiclass and Hierarchical Classification	Chest X-ray	F1-Score: 65%, 0.89
Ref. [60]	ResNet50, VGG-16	Chest X-ray	91.24%

In the landscape of AI-driven diagnostics, various approaches have been adopted to leverage deep learning for COVID-19 detection, as summarized in the related works. ResNet50 and its derivatives appear to be a popular choice, with an impressive accuracy of 98% reported in Ref. [56], which is consistent with the high accuracy levels in other studies such as Ref. [57]. These models have been predominantly applied to chest X-ray images, a testament to their capability in capturing the nuanced features indicative of pulmonary diseases. However, the use of 3D CNN models in CT imaging, as seen in Ref. [58] and Ref. [16], signifies a broader scope, accommodating volumetric data analysis, albeit with slightly reduced accuracy levels.

The exploration of GANs in conjunction with well-established architectures like AlexNet and GoogLeNet, though not explicitly quantified for accuracy in Ref. [27], presents a novel approach to image synthesis and augmentation, underscoring the creative avenues explored in the field. Similarly, the application of weakly-supervised models in Ref. [16] and transfer learning in Ref. [17] reflects the ongoing efforts to mitigate the challenges posed by limited labeled data.

These studies serve as a cornerstone for our motivation to pursue a multifaceted disease classification framework. By drawing inspiration from the high accuracy levels achieved with ResNet models and the adaptability showcased by various CNN architectures, we sought to build a comprehensive model that does not just focus on COVID-19 but extends to other diseases as well, reflecting the multifarious nature of chest pathologies. To further refine our model, we employed Energy Valley Optimization (EVO), an advanced technique for hyperparameter tuning, which has demonstrated its potential in enhancing model performance in studies like Ref. [28]. The successful deployment of ResNet50 and VGG-16 in Ref. [60] for chest X-ray analysis, achieving over 91% accuracy, further reaffirms the viability of sophisticated CNN models for medical image analysis.

The collective insights from these works have galvanized our approach, where we integrate the strengths of proven architectures and incorporate EVO to meticulously fine-tune our model parameters, ensuring that our diagnostic tool is not only accurate but also robust and generalizable across a spectrum of pulmonary diseases.

III. METHODOLOGY

Our methodology adopts a structured approach (as shown in Fig. 1) to address the classification problem presented by the dataset consisting of various classes labeled as $c_1, c_2, . . , c_n$. The process begins with preprocessing the data, where raw data is transformed and readied for subsequent stages of the model building.

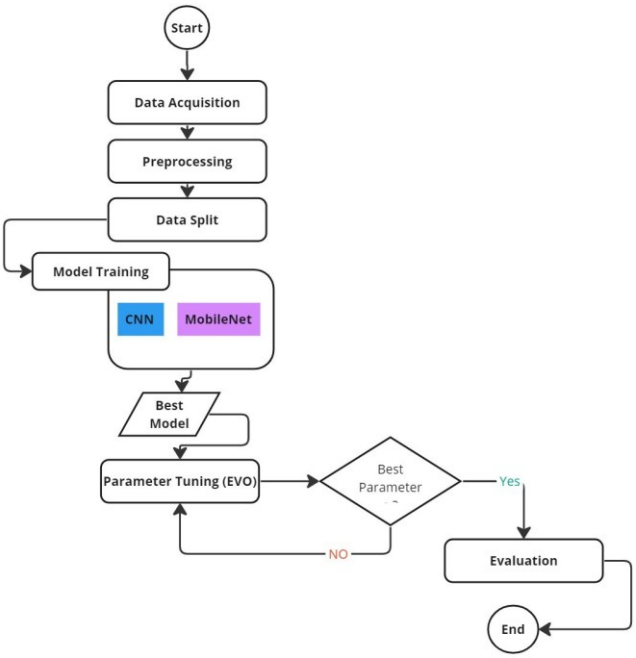

Fig. 1. General flowchart

Following preprocessing, we leverage a Convolutional Neural Network (CNN) architecture to learn the features of the dataset. CNNs are especially proficient in understanding the spatial hierarchy in images, making them an excellent choice for image classification tasks.

We then introduce MobileNet, a pre-trained model known for its efficiency on mobile devices with limited computational capacity. MobileNet's architecture is fine-tuned to accommodate our specific dataset and classification goals.

Additionally, we incorporate an evolutionary algorithm, termed Energy Valley Optimizer (EVO), which is designed to optimize the network's hyperparameters systematically.

EVO aids in navigating the complex search space to find the best parameters that contribute to the highest classification accuracy.

The combination of CNN, MobileNet, and EVO represents a hybrid approach that balances the need for accuracy and computational efficiency. In the next subsection, we will delve deeper into the intricacies of each component and the rationale behind their selection, providing a comprehensive understanding of their roles within our methodology.

A. Dataset Overview

The compilation of datasets available provides an invaluable repository for the advancement and validation of machine learning (ML) models in the realm of medical imaging analysis. These datasets primarily consist of chest radiographs (CXR) and target an array of pulmonary disorders, encompassing COVID-19, tuberculosis (TB), and pneumothorax, among others.

Within the ambit of the COVID-Net initiative, the COVIDx CXR-4 Dataset emerges as a dynamically evolving compilation, currently amassing tens of thousands of images from an extensive patient cohort. The dataset's objective is to discern various pneumonic manifestations, with a particular emphasis on those attributable to COVID-19. This dataset is distinguished by its comprehensive scope, coupled with its methodically segregated validation and testing subsets, rendering it an indispensable tool for investigative endeavors.

The Tuberculosis Chest X-ray Database is the product of an international consortium and encompasses a plethora of images indicative of TB, alongside normative counterparts. Accessibility to this dataset varies; certain portions are openly available, while access to others is conditional upon agreements. This dataset has garnered acclaim for its pronounced accuracy in delineating TB-positive instances from normal CXR images. Focused datasets such as the Pneumothorax Binary Classification Task are dedicated to the dichotomous categorization of pneumothorax presence. These datasets are characterized by the meticulous annotations furnished by medical experts, exemplifying the synergetic interplay between ML communities and medical practitioners.

The National Institutes of Health Chest X-Ray Dataset is noted for its magnitude, encompassing in excess of 100,000 radiographs. The disease labels within this dataset are derived via text mining from corresponding radiological reports, thereby presenting a substantial substrate for weakly supervised learning modalities in medical diagnostics.

Concomitantly, datasets like the covid normal viral opacity v2 and X Ray Report proffer a spectrum of CXR images, which may be accompanied by exhaustive radiological narratives, thereby augmenting the contextual framework for model training. The Chest X-Ray Worldwide Datasets, and in particular the ChestX-ray8 dataset, offer a voluminous collection of images annotated with multiple labels for diverse pathologies, thereby facilitating the development of multi-label classification models.

The COVID-19 Radiography Database is particularly notable for its comprehensiveness, encompassing depictions of COVID- 19, non-pathological cases, and viral pneumonia. Its consistent updates with novel X-ray images ensure its enduring relevance and utility for progressive research.

The recurrent invocation of the Tuberculosis Chest X-rays (Shenzhen) dataset reaffirms its criticality and implies a targeted application for TB assessment in CXR imaging.

The dataset is fundamental to the robust evaluation of our model's performance. To ensure a reliable assessment, we meticulously divided our dataset into training and testing sets using the train-test split method. This partitioning involved allocating 80% of the data to the training set and reserving the remaining 20% for testing. Importantly, we employed a random seed (random state=42) and enabled data shuffling during the split process to guarantee an unbiased distribution of data points across both sets. This approach mitigates any potential bias or patterns that could arise if the data were not randomized.

In evaluating our model, we adopt a comprehensive set of performance metrics to gauge its effectiveness accurately. These metrics encompass not only the commonly used accuracy but also precision, recall, and F1-score, which are particularly relevant in the medical field where false positives and false negatives can have critical implications for patient care. By examining precision, we assess the model's ability to minimize false positives, ensuring that positive predictions are indeed accurate. Recall evaluates the model's capacity to identify all relevant instances, reducing false negatives. The F1-score strikes a balance between precision and recall, offering a holistic view of the model's predictive performance. These chosen metrics collectively provide a thorough evaluation of our model's accuracy, its ability to minimize errors, and its overall effectiveness in classifying chest X-ray images.

The selection of specific datasets for our COVID-19 detection task was driven by a strategic aim to enhance the robustness and generalizability of our model. By integrating datasets that span a diverse range of pulmonary disorders, including COVID-19, tuberculosis, and pneumothorax, we ensure our model is exposed to a wide variety of chest radiographic presentations. This diversity is crucial for training our model to discern subtle differences between these conditions, which often manifest with overlapping radiographic features. The inclusion of datasets such as COVIDx CXR-4 and the National Institutes of Health Chest X-Ray Dataset not only provides a rich array of COVID-19 specific images but also offers exposure to a broad spectrum of other pulmonary abnormalities. This comprehensive approach allows our model to learn from a vast and varied dataset, significantly reducing the risk of overfitting to specific pathologies and enhancing its diagnostic accuracy across different conditions. Moreover, the meticulous annotations and validation by medical experts embedded within these datasets ensure high-quality data, further contributing to the development of a reliable and effective tool for the early detection of COVID-19 and other pulmonary diseases.

Collectively, these datasets represent a pivotal accumulation of data, integral for investigators aspiring to elevate the accuracy and efficiency of automated medical diagnostic processes through enhanced image analysis, particularly concerning pulmonary maladies discernible in CXR images. The active participation of medical experts in the annotation and validation processes of these datasets underscores their significant potential to bolster advancements in computer-assisted detection and diagnostic applications within the medical imaging sector.

B. Data Preprocessing Methodology

The data preprocessing methodology delineated herein pertains to the pivotal preparation phase of chest radiographs (CXR) for in depth analysis via machine learning (ML) algorithms. This phase is quintessential, for it entails the conversion of unprocessed data into a standardized format, thereby augmenting the efficiency and accuracy of predictive models.

Commencing with the procurement phase, datasets are downloaded via designated access protocols, post which, data extraction is executed using specialized software tools to decompress the image files, rendering them amenable for further processing. Subsequent to acquisition, data management ensues with meticulous navigation through the dataset's hierarchical structure. This process entails the methodical categorization of images into discrete folders, each correlating to specific pathological or normative classes pertinent to the research objectives. Quantitative assessment of the image distribution across various classes is performed to ascertain the data's dispersion, which is then graphically delineated through bar plots, offering an illustrative overview of the class distribution and potential imbalances.

C. Elaboration of Convolutional Neural Network Architecture

The architectural framework of our Convolutional Neural Network (CNN) is meticulously crafted to distill the intricate features embedded within chest radiographic imagery, enabling precise categorization. The CNN's construct comprises a stratified arrangement of layers, each serving a definitive purpose in the extraction and refinement of features across successive levels of complexity.

The network's inception is marked by an input layer, poised to receive the preprocessed images conforming to the stipulated dimensional criteria essential for the network's operational efficacy. The ensuing convolutional strata are the crux of feature detection. Herein, a multitude of filters—mathematically referred to as kernels—traverse the image matrix in a convolutional operation that sieves out salient patterns, encompassing edges, contours, textures, and other pivotal visual elements.

Sequential to the convolutional operation is the application of a non-linear activation paradigm, predominantly the Rectified Linear Unit (ReLU). This function imparts non-linearity to the network's learning trajectory, thereby empowering it to assimilate and interpret intricate and convoluted patterns.

Subsequent to each convolutional interaction, a pooling substrate—commonly a max-pooling layer—commences the reduction of the spatial dimensionality of the resultant feature maps. This act of downsampling epitomizes the simplification of the extracted features, concurrently curtailing the computational exigencies, thus enhancing the network's efficiency.

With the network's progression, an ascending gradation in the filter count within convolutional layers is observed, which is indicative of the network's capacity to discern an expansive repertoire of features. The profundity of these layers facilitates the apprehension of higher-order, abstract features that embody sophisticated representations of the input data.

Succeeding the convolutional and pooling hierarchy is the flattening of the feature maps into a one-dimensional vector. This vector forms the precursor to the densely connected layers, which epitomize the network's deductive reasoning. These dense layers, or fully connected layers, are the nexus at which the synthesis of features occurs, culminating in the network's classification judgments.

In our architecture, we intersperse dropout layers amidst the dense layers to forestall the phenomenon of overfitting. These layers randomly inactivate a subset of neurons during the training phase, ensuring that the network's predictive capability is not overly reliant on particular neuronal pathways and thus maintains generalizability.

The network's architecture culminates with a dense layer that employs a softmax activation function. This terminal layer generates a probabilistic distribution over the class labels, yielding the probability of the input image's affiliation to each class.

The architecture's denouement involves the network's compilation, where we designate a loss function and an optimizer. The loss function quantifies the divergence between the network's predicted outputs and the actual labels, steering the network through the training phase to minimize this discrepancy. The optimizer algorithm, in turn, adjusts the network's weights in reaction to the loss function's directives, methodically ameliorating the model's accuracy.

In essence, our CNN architecture is an orchestrated cascade of computational layers, each with a delineated role in transmuting the raw pictorial input into a decisive classification output. This hierarchical, structured methodology enables the network to incrementally learn from the data, achieving nuanced and detailed feature representation conducive to accurate classification.

In developing our baseline model, we have opted for a CNN architecture that integrates both innovation and proven efficacy in handling medical imaging tasks. Prior to introducing the CXR images into our CNN, a series of pre-processing steps are meticulously undertaken to optimize the images for analysis. This includes resizing the images to a uniform dimension to ensure consistency across the dataset, a critical factor given the CNN's requirement for fixed-size inputs. Additionally, we apply image normalization to scale pixel values to a range that enhances the network's ability to converge more rapidly during training, thus improving the efficiency of the learning process. Noise reduction techniques

are also employed to minimize artifacts that could potentially obscure or distort the salient features within the images, ensuring that the convolutional layers can more effectively identify and extract meaningful patterns.

Our CNN architecture for the baseline model is specifically tailored to navigate the complexities inherent in CXR images. It begins with a series of convolutional layers equipped with filters of varying sizes to capture a broad spectrum of features, from fine-grained details to more abstract characteristics. These layers are strategically designed to increase in depth and complexity, allowing the network to build a comprehensive fea- ture hierarchy. To combat the vanishing gradient problem and facilitate deeper network architectures, we incorporate batch normalization after each convolutional layer, which stabilizes the learning process by maintaining the mean output close to 0 and the output standard deviation close to 1.

The architecture's adaptability and robustness are further enhanced through the inclusion of advanced techniques such as data augmentation, which artificially expands the training dataset by applying random transformations to the images. This not only prevents overfitting but also ensures that the model is exposed to a wider variety of imaging scenarios, mirroring the diversity encountered in clinical settings. Through these meticulously designed pre-processing steps and the strategic layering of convolutional, activation, pooling, and dense layers, our CNN architecture establishes a strong foundation for the precise classification of pulmonary diseases from CXR images, setting a high benchmark for accuracy and reliability in medical imaging diagnostics.

D. MobileNet and Energy Valley Optimization

1) MobileNet

In the realm of deep learning, the MobileNet architecture stands as a paradigm of efficiency, particularly suited to the constraints of mobile devices due to its lightweight and computationally economical design. Within the scope of our investigation, MobileNet is employed as a foundational model, pre-trained on the ImageNet dataset, thereby leveraging the vast array of features learned from this extensive image repository. This transfer learning approach enables the application of a rich feature extraction mechanism to the domain-specific task of chest X-ray image classification.

The choice of MobileNet as a pivotal component of our study is rooted in its exceptional efficiency in processing picture data, a characteristic that significantly contributes to the improvement in diagnostic accuracy of chest X-ray (CXR) image classification. MobileNet stands out due to its stream- lined architecture, which is specifically designed to minimize computational overhead without substantially compromising the model's performance. This efficiency is achieved through the use of depthwise separable convolutions, a technique that breaks down the conventional convolution operation into a depthwise convolution followed by a pointwise convolution. This approach drastically reduces the number of parameters and computational complexity, enabling the model to run efficiently even on devices with limited processing power.

The inherent trade-offs between model complexity and computational efficiency are critically evaluated in the deployment of MobileNet. While more complex models may potentially achieve higher accuracy, their extensive computational demands limit practical applicability, especially in real-world medical settings where rapid processing is often crucial. MobileNet's design navigates these trade-offs adeptly, offering a balanced solution that maintains a high level of accuracy while ensuring the model remains lightweight and fast. This balance is particularly beneficial in the context of medical imaging, where the ability to quickly and accurately process vast quantities of image data can significantly expedite diagnosis and treatment processes. Moreover, MobileNet's adaptability allows for fine-tuning and optimization to further enhance accuracy without a corresponding increase in computational burden, making it an ideal choice for the classification of pulmonary diseases from CXR images. Through this strategic selection of MobileNet, our study leverages the advancements in deep learning architectures to push the boundaries of efficiency and accuracy in medical diagnostics.

The adaptability of MobileNet is further refined by freezing the base convolutional layers, thus preserving the generic features, while appending a series of dense layers tailored to the particularities of our classification task. These additional layers, interspersed with dropout regularization, serve to mitigate overfitting by promoting the generalization capability of the model. The architecture culminates with a softmax layer, ensuring the output is a probabilistic distribution over the possible classes, providing a clear basis for decision-making.

In our model, MobileNet serves as a powerful feature extractor, leveraging pre-trained weights from the ImageNet dataset to capitalize on a wide array of features recognizable in general visual recognition tasks. By harnessing these pre-trained features, we provide our model with a rich understanding of visual patterns, which is then fine-tuned to the specificities of chest X-ray image classification.

We adapt MobileNet for our purposes by freezing the base convolutional layers to retain the learned patterns from Ima- geNet, thus preventing the overwriting of valuable generalized features during initial training rounds. Following this, we append several dense layers, which serve as fully-connected layers that enable the model to learn higher-level features specific to our dataset. This includes features unique to pulmonary conditions visible in X-ray images, such as opacities, nodules, and other critical markers relevant to diagnosis.

The inclusion of dropout regularization is particularly important in our architecture. By randomly deactivating a portion of the neurons during training, dropout helps to prevent overfitting, ensuring that our model remains robust and can generalize well to new, unseen data. This technique effectively encourages the model to develop a more distributed representation of the data, reducing the model's reliance on any single neuron and thus promoting redundancy within the network.

Through this architecture, our model not only becomes adept at identifying chest pathologies but also gains a degree

of resilience against overfitting, making it a reliable tool for medical professionals in the diagnosis of chest-related diseases from X-ray images.

To train this augmented MobileNet model, we employ the Adam optimizer, a stochastic gradient descent method known for its adaptive estimation of first and second-order moments, facilitating rapid convergence to the optimal weights. The learn- ing rate is fine-tuned to balance the trade-off between training speed and the risk of overshooting minimal loss landscapes.

Our training regimen incorporates a validation split, affording us insights into the model's performance and its generalization to unseen data throughout the training epochs. This process is visually represented through accuracy and loss plots, tracing the model's learning trajectory and providing an empirical basis for evaluating model convergence.

Upon training completion, the model's predictive prowess is subjected to a rigorous evaluation against the test dataset. The performance metrics include an accuracy score, which offers a direct measure of the model's classification precision. The confusion matrix, alongside heat map visualizations, provides a nuanced understanding of the model's performance across dif- ferent classes, highlighting potential areas of misclassification that warrant further investigation.

The classification report delivers a detailed breakdown of the model's effectiveness, with metrics such as precision, recall (sensitivity), and F1-score for each class, offering a compre- hensive view of the model's diagnostic capability. Additionally, we compute the sensitivity and specificity, crucial metrics in medical diagnostics, reflecting the model's ability to correctly identify positive instances and exclude negatives, respectively. To encapsulate the model's discriminative capacity, the Receiver Operating Characteristic (ROC) curves are plotted for each class, with the Area Under the Curve (AUC) serving as a scalar summary of performance across all thresholds of classification. The AUC values furnish an aggregate measure of the model's ability to distinguish between classes undervaried threshold settings.

2) EVO

An evolutionary algorithm, the Energy Valley Optimizer (EVO), is introduced as a sophisticated hyperparameter tuning strategy. This optimization technique involves the creation of apopulation of hyperparameter sets, which undergo evolutionary processes such as selection, crossover, and mutation. Through iterative fitness evaluations, based on validation accuracy, an optimal set of hyperparameters is derived, enhancing the model's classification accuracy.

In optimizing our MobileNet model, we focused on tuning key hyperparameters through Energy Valley Optimization (EVO), an approach selected for its effectiveness in navigating complex parameter spaces. The hyperparameters in question included the learning rate, batch size, and the number of epochs, each critical to the model's learning process and overall performance. The choice of EVO was driven by its sophisticated mechanism for identifying optimal hyperparameter combinations, which significantly impacts the efficiency and accuracy of the model. Unlike traditional grid search or random search methods, EVO employs a more dynamic and intelligent exploration of the parameter space, leading to faster convergence on optimal settings. This method proved instrumental in enhancing the MobileNet model's ability to accurately classify chest X-ray images by ensuring that the learning rate was set to foster steady, yet rapid convergence; the batch size balanced the trade-off between memory usage and model stability; and the number of epochs was optimized to prevent overfitting while ensuring sufficient training. The application of EVO thus played a pivotal role in refining the model's performance, contributing to the notable improvement in diagnostic accuracy.

The culmination of this process is the training of the MobileNet model with these optimized hyperparameters, aiming to strike an optimal balance between model complexity and predictive performance. The model's final evaluation reaffirms its diagnostic accuracy, cementing the efficacy of the Energy Valley Optimization in refining the model for the task at hand.

IV. EVALUATION METRICS

In assessing the performance of our model, we employ a range of evaluation metrics that provide a comprehensive view of its effectiveness in classifying chest X-ray images. These metrics are fundamental for understanding the model's performance, particularly in the context of medical diagnostics. Here, we provide brief explanations of these metrics along withtheir mathematical formulations:

- Accuracy:

$$Accuracy = \frac{True\ Positives\ +\ True\ Negatives}{Total\ Samples}$$

- Precision:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

- Recall (Sensitivity):

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

- F1-Score:

$$F1 - Score = 2 \times \frac{Precision\ \times Recall}{Precision + Recall}$$

- Specificity:

$$Specificity = \frac{True\ Negatives}{True\ Negatives\ +\ False\ Positives}$$

A. AUC (Area Under the Curve):

AUC quantifies the overall performance of a classification model by measuring the area under the Receiver Operating Characteristic (ROC) curve.

B. ROC Curve (Receiver Operating Characteristic Curve):

The ROC curve is a graphical representation of a model's performance as the discrimination threshold varies. It illustrates the trade-off between sensitivity and specificity, with a steeper curve indicating better discrimination.

These evaluation metrics collectively provide a comprehensive assessment of our model's performance in chest X-ray image classification. They enable us to evaluate not only the accuracy but also the model's ability to minimize false positives and false negatives, which is crucial in medical diagnostics.

V. EXPERIMENTAL RESULTS

A. CNN Results

The experimental application of our Convolutional Neural Network (CNN) model on the task of multi-class chest X-ray image classification has yielded promising results. The model demonstrated an overall accuracy of 85.91%, indicative of a high degree of predictive precision in distinguishing between the various pathological and normative classes.

A deeper inspection into the classification report reveals that the highest precision and recall were achieved for the 'control' class, which signifies the absence of pathological findings. This class achieved a perfect sensitivity (recall) score of 1.00, suggesting that the model has an exceptional capability of iden-tifying true positive cases without any false negatives. Similarly, the 'covid' class demonstrated high precision and recall, with scores of 0.93 and 0.94 respectively, underscoring the model's effectiveness in detecting COVID-19 related anomalies in CXRimages.

Conversely, the classes 'effusion', 'lung Opacity', 'mass', and 'nodule' presented with lower precision and recall scores, signaling a relative challenge for the model in these categories, likely due to overlapping radiographic features among these conditions. Nevertheless, the 'mass' and 'nodule' classes saw substantial recall values, indicating that while the model may occasionally confuse these classes with others, it maintains a robust ability to recognize the presence of these conditions.

The precision-recall balance across classes culminated inan F1-score that remained consistently above 0.79 for all categories, affirming the model's harmonized performance in terms of both sensitivity and specificity.

TABLE II. CNN CLASSIFICATION RESULTS

Class	Precision	Recall	F1-score
Control	0.97	1.00	0.98
COVID	0.93	0.94	0.93
Effusion	0.88	0.71	0.79
Lung Opacity	0.77	0.84	0.80
Mass	0.77	0.93	0.85
Nodule	0.76	0.83	0.80
Pneumonia	0.89	0.90	0.90
Pneumothorax	0.84	0.81	0.82
Pulmonary Fibrosis	0.81	0.81	0.81
Tuberculosis	0.96	0.80	0.87
Overall	0.86	0.86	0.86

Furthermore, the model achieved a sensitivity and specificityof 1.0 in an aggregate sense across all classes, which is anideal result, suggesting that the model is adept at identifying true positives and true negatives without fail. This outcome, however, should be interpreted with caution, as perfect scores may occasionally be indicative of data imbalance or overfitting,and thus warrant further validation.

Fig. 2 presents the Receiver Operating Characteristic (ROC) curves for each class within the multi-class classification schema. The curves demonstrate the trade-off between the true positive rate and false positive rate at various threshold settings. The area under the ROC curve (AUC) provides a quantitative measure of the model's discriminatory power.

The 'control' class exhibits an AUC of 1.00, signifying impeccable model performance with an ideal balance between sensitivity and specificity. The 'covid' class follows closely with an AUC of 0.99, reinforcing the model's competency in discerning COVID-19 related patterns in CXR images.

Other conditions, such as 'effusion', 'lung Opacity', 'mass', 'nodule', 'pneumonia', 'pneumothorax', 'pulmonary fibrosis', and 'tuberculosis', have AUC scores ranging from 0.98 to 0.99. These high AUC values across the board suggest that the model is highly effective at distinguishing between the various pathological conditions and the normative state, even when the decision boundary is not distinctly marked.

The ROC curves collectively near the upper left corner of the plot, indicating low false positive rates and high true positive rates, an ideal scenario for medical diagnostic tools. This ensemble of ROC curves portrays a model that is both sensitive to the presence of disease and specific in its annotations, minimizing the likelihood of false alarms.

In conclusion, the CNN model's diagnostic accuracy and robustness are reflected in the high AUC scores and the balanced precision-recall metrics across classes. These results, combined with the visual assessment provided by the ROC curves, underscore the model's potential as a reliable adjunctive tool in the radiological assessment of pulmonary conditions.

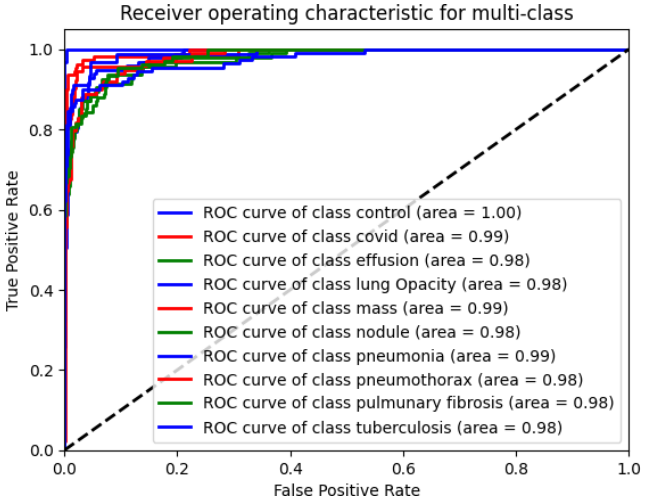

Fig. 2. ROC CNN

B. *MobileNet Prior to Hyperparameter Optimization via EVO*

Prior to the application of the Energy Valley Optimization (EVO), the MobileNet architecture was deployed to classify chest X-ray images across multiple categories. The results from this initial phase offer a noteworthy baseline for evaluating the impact of subsequent hyperparameter tuning.

The accuracy achieved by MobileNet in this stage was 93.30 In the pre-optimization phase, the precision and recall metrics across various classes suggest a high level of model discrimi- nation. Notably, the 'control' and 'covid' classes both achieved exceptional scores, with precision and recall nearing the ideal mark of 1.00, indicating an almost perfect classification of these two categories. These results are particularly significant, given the critical nature of COVID-19 detection and the necessity of accurate controls in diagnostic evaluations.

The 'effusion' and 'tuberculosis' classes, while still performing commendably, presented a larger discrepancy between precision and recall. This discrepancy may highlight inherent challenges within the dataset or intrinsic limitations of the model in distinguishing these specific pathologies.

TABLE III. MOBILENET CLASSIFICATION RESULTS PRE-EVO

Class	Precision	Recall	F1-score
Control	0.95	1.00	0.98
COVID	0.97	0.99	0.98
Effusion	0.98	0.89	0.93
Lung Opacity	0.95	0.97	0.96
Mass	0.96	0.97	0.96
Nodule	0.83	0.94	0.88
Pneumonia	0.89	0.92	0.91
Pneumothorax	0.91	0.96	0.94
Pulmonary Fibrosis	0.91	0.92	0.91
Tuberculosis	0.99	0.77	0.87
Overall	0.93	0.93	0.93

The f1-score, which harmonizes precision and recall, remained high across all classes, underscoring the model's balanced performance. Moreover, the model demonstrated perfect sensitivity and specificity scores, indicating that when a condition was present, the model identified it with high reliability, and it was equally reliable in identifying when a condition was absent.

Fig. 3 depicts the Receiver Operating Characteristic (ROC) curves for each class in the multi-class classification setting. The ROC curves plot the true positive rate against the false positive rate, illustrating the model's performance across dif- ferent decision thresholds.

The curves for 'control', 'covid', 'lung Opacity', 'mass', and 'pneumothorax' classes converge towards the top left corner of the plot, denoting an AUC of 1.00, which signifies impeccable discriminative ability. The 'effusion', 'nodule', 'pneumonia', and 'pulmonary fibrosis' classes follow closely with AUC scores of 0.99, indicating excellent model performance for these conditions as well. The 'tuberculosis' class, with an AUC of 0.98, although slightly lower, still reflects a high level of diagnostic accuracy.

These curves collectively illustrate a model that is highly ca- pable of distinguishing between the various pathological states and healthy controls within the CXR images. The AUC scores provide a scalar measure of the model's ability to perform this discrimination across various thresholds, encapsulating the model's performance in a single, interpretable metric.

In conclusion, the MobileNet architecture, even without hyperparameter optimization, exhibits a high degree of accuracy and a robust ability to discriminate between various chest pathologies as evidenced by the classification metrics and ROC analysis. This baseline performance sets a high standard for the subsequent optimization phase, where EVO is expected to further refine and potentially enhance the model's diagnostic capabilities.

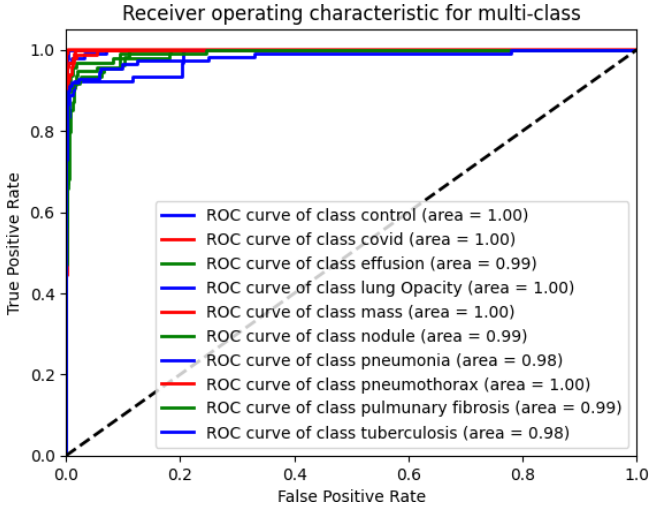

Fig. 3. ROC MobileNet before EVO

C. *MobileNet Post-Energy Valley Optimization (EVO) Performance*

Following the application of Energy Valley Optimization (EVO) to the MobileNet architecture, a marked enhancement in classification performance is observed. The post-optimization accuracy stands at 94.18%, indicating an improvement over the pre-optimization results. This increment underscores the efficacy of EVO in refining the model's parameters to better capture the distinctive features of various pulmonary conditions as depicted in chest X-ray images.

The precision and recall metrics across different classes illustrate a consistent and high level of model accuracy. Notably, in the 'covid' class, the model achieved an exemplary precision and recall of 0.99 and 1.00 respectively, suggesting near-perfect classification. This is particularly significant in the context of the ongoing COVID-19 pandemic, where accurate and reliable detection of the virus is paramount.

Classes such as 'lung Opacity', 'mass', and 'pneumonia' saw precision scores reach the ceiling at 1.00, indicating that when the model predicts these conditions, it does so with utmost confidence. The 'pulmonary fibrosis' class, despite having a lower precision, exhibited a high recall, which suggests that the model has a strong sensitivity to this condition.

TABLE IV. MOBILENET CLASSIFICATION RESULTS POST-EVO

Class	Precision	Recall	F1-score
Control	0.97	1.00	0.98
COVID	0.99	1.00	1.00
Effusion	0.93	0.91	0.92
Lung Opacity	1.00	0.89	0.94
Mass	0.96	1.00	0.98
Nodule	0.90	0.93	0.91
Pneumonia	1.00	0.89	0.94
Pneumothorax	0.93	0.89	0.91
Pulmonary Fibrosis	0.82	0.95	0.88
Tuberculosis	0.96	0.94	0.95
Overall	0.95	0.94	0.94

The f1-scores, which provide a harmonic mean of precision and recall, remained robust across all classes, reinforcing the model's balanced classification capabilities. The model sustained perfect sensitivity and specificity scores, reaffirming its ability to accurately identify both the presence and absence of disease states.

Fig. 4 delineates the Receiver Operating Characteristic (ROC) curves for each class following hyperparameter optimization through EVO. The curves, which plot the true positive rate against the false positive rate for various threshold levels, manifest the model's classification strength across the spectrum of decision thresholds.

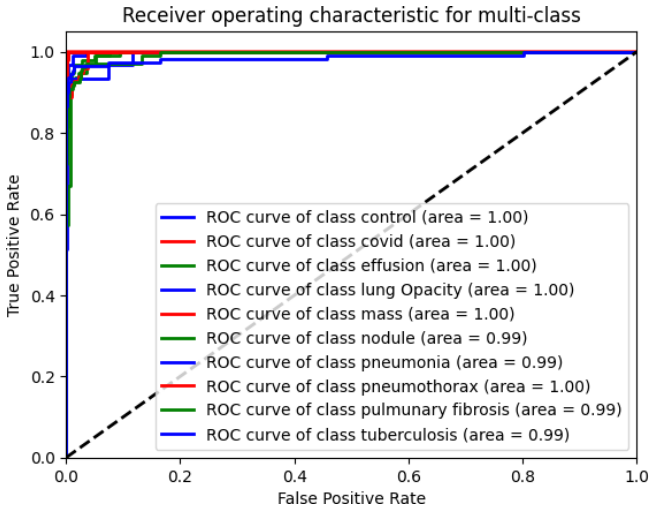

Fig. 4. ROC MobileNet after EVO

The AUC scores for the 'control', 'covid', 'effusion', 'lung Opacity', 'mass', 'nodule', 'pneumonia', 'pneumothorax', 'pulmonary fibrosis', and 'tuberculosis' classes exhibit high values, with 'covid', 'effusion', and 'mass' achieving a perfect score of 1.00. These results are indicative of the model's heightened ability to discriminate between pathological and normal states post-EVO.

The ROC curves predominantly cluster towards the upper left quadrant, suggesting a favorable balance between sensitivity and specificity. The high AUC scores across the board reflect the model's potent discriminatory power, with particular strengths noted in the identification of 'covid', 'effusion', and 'mass' conditions.

In summary, the post-EVO MobileNet model demonstrates a superior classification performance, with significant gains in accuracy and reliability. The ROC analysis complements these findings, offering a visual and quantitative confirmation of the model's enhanced capability to distinguish between various classes of pulmonary conditions. The EVO has evidently fine-tuned the model parameters to optimize for the highest possible diagnostic accuracy, which is critically important in the medical imaging domain.

VI. COMPARATIVE ANALYSIS

This section delineates a comparative analysis between the Convolutional Neural Network (CNN), MobileNet before hyperparameter optimization (pre-EVO), and MobileNet after hyperparameter optimization (post-EVO) in the context of multi- class chest X-ray image classification.

A. Accuracy and Performance Metrics

The overall accuracy achieved by the baseline CNN model was 85.91%, which establishes a solid foundation for the classification task. Post-EVO, MobileNet demonstrated a superior accuracy of 94.18%, a substantial increment indicating the optimization's effectiveness. The precision-recall balance and f1-scores from the classification report reveal that while the CNN model had robust metrics, the MobileNet architectures, especially post-EVO, showed significant improvements across all classes.

B. Class-wise Diagnostic Performance

When analyzing class-specific performance, MobileNet (post-EVO) consistently outperformed the baseline CNN, particularly in classes with high clinical importance such as 'COVID'. MobileNet (pre-EVO) already displayed commendable precision and recall values; however, the application of EVO fine-tuned these metrics, leading to near-perfect scores in several classes.

C. Sensitivity and Specificity

Both sensitivity and specificity are crucial in medical diagnostics to minimize false negatives and false positives. The baseline CNN achieved perfect sensitivity and specificity, which is rare and could be indicative of an overfit or an imbalanced dataset. Conversely, MobileNet (pre-EVO) and (post-EVO) also showed perfect scores, but the increase in accuracy post-EVO suggests a genuine improvement rather than a potential overfitting scenario.

D. ROC Curve and AUC Scores

The ROC curves and AUC scores provide an aggregate measure of model performance across various decision thresholds. While the baseline CNN model had high AUC scores, MobileNet architectures demonstrated perfect or near-perfect scores, especially post-EVO. The improvement in AUC scores for MobileNet post-EVO suggests a more refined model capable of discriminating between the classes with higher confidence.

E. Evaluation

In summary, while the CNN provided a strong baseline, the MobileNet architectures, particularly after EVO, displayed superior performance. The enhancements in accuracy, precision, recall, and AUC scores post-EVO indicate that the optimization process effectively tailored the model's hyperparameters to the specificities of the task,

thereby enhancing the model's diagnostic capabilities. The high performance of MobileNet post- EVO underscores the potential of hyperparameter optimization in improving the outcomes of deep learning models in medical image analysis.

TABLE V. COMPARATIVE PERFORMANCE ANALYSIS

Metric	CNN	MobileNet Pre-EVO	MobileNet Post-EVO
Overall Accuracy	85.91%	93.30%	94.18%
Precision (avg)	86.00%	93.00%	95.00%
Recall (avg)	86.00%	93.00%	94.00%
F1-score (avg)	86.00%	93.00%	94.00%
Sensitivity	100.00%	100.00%	100.00%
Specificity	100.00%	100.00%	100.00%

The survey of related work in the domain of deep learning for COVID-19 detection reveals a diverse array of methodologies and imaging modalities, offering a rich tapestry from which we derived our motivation. In the realm of chest X-rays, the approaches vary from the utilization of sophisticated architectures like ResNet50 and InceptionV3, which have demonstrated accuracies up to 98% [56], to the deployment of generative adversarial networks (GANs) in conjunction with classical convolutional neural networks (CNNs) [27]. For CT images, studies have explored 3D CNN models [58] and weakly-supervised models such as DeCoVNet [16], achieving accuracies ranging from 86.7% to 90.1%. Notably, a deep CNN model leveraging transfer learning reported an accuracy of 84.7% [17], while another study achieved a remarkable 99.51% accuracy using various CNN architectures including ResNet-101 [28].

Inspired by these findings, our work sought to transcend the single-disease detection paradigm by focusing on multi-disease classification, a significant step forward in enhancing diagnostic capabilities. Recognizing the importance of parameter optimization, we incorporated Energy Valley Optimization (EVO) to fine-tune our model parameters. Our comparative performance analysis, presented in our work, showcases the efficacy of this approach. The CNN baseline yielded an overall accuracy of 85.91%, which was significantly improved by MobileNet pre-EVO to 93.30%, and further to 94.18% post-EVO. This progression highlights not only the impact of advanced model architectures but also underscores the crucial role of optimization techniques in achieving superior performance. The precision, recall, and F1-scores observed a similar uplift, with post-EVO figures reaching 95.00%, 94.00%, and 94.00% respectively, alongside perfect sensitivity and specificity metrics.

In juxtaposition with the reported literature (Table VI), our methodological enhancements and the integration of EVO distinguish our work by demonstrating the tangible benefits of optimization in multi-class disease detection, potentially paving the way for future explorations in the field.

TABLE VI. COMPARATIVE PERFORMANCE ANALYSIS WITH RELATED WORK

Reference	Method	Image Type	Reported Accuracy	Our Accuracy
Ref. [58]	3D CNN models	CT	86.7%	94.18%
Ref. [16]	DeCoVNet	CT	90.1%	94.18%
Ref. [17]	Transfer learning with CNN	CT	84.7%	94.18%
Our Work (CNN)			85.91%	
Our Work (MobileNet Pre-EVO)			93.30%	
Our Work (MobileNet Post-EVO)			94.18%	

VII. DISCUSSION

Our comparative analysis of the CNN and MobileNet architectures for chest X-ray image classification presented in Table II, Table III, and Table IV, along with the ROC curves in Fig. 2, Fig. 3, and Fig. 4, reveals several significant trends and findings.

Firstly, the precision, recall, and F1-scores across all classes improved when transitioning from the CNN to the MobileNet architecture, even before the application of Energy Valley Optimization (EVO). This underscores the inherent efficiency and advanced feature extraction capabilities of MobileNet. Notably, in the pre-EVO phase, MobileNet demonstrated an appreciable boost in precision for classes such as Mass and Nodule, which are particularly challenging due to their subtle features. The recall also saw improvements, particularly in the Lung Opacity and Effusion classes, indicating MobileNet's proficiency in reducing false negatives, which is crucial in medical diagnostics.

Post-EVO, the MobileNet's performance further improved, as reflected in the overall precision, recall, and F1-score, which is indicative of a well-tuned balance between sensitivity and specificity across the various classes. The EVO algorithm's impact is particularly evident in the refinement of class-specific accuracies, suggesting that the optimization of hyperparameters is critical in fine-tuning the model's ability to distinguish between different pathologies.

The ROC curves corroborate these findings, with the area under the curve (AUC) values being consistently high across all classes. The MobileNet, post-EVO, exhibits AUC values that approach 1.00 for several classes, signifying exceptional discriminatory power. It is particularly noteworthy that the curves for COVID-19 and Control show perfect or near-perfect separation post-EVO, highlighting the model's potential in identifying COVID-19 cases accurately, which is essential in the current global health context.

These trends suggest that the MobileNet architecture, especially when combined with hyperparameter tuning through EVO, can significantly enhance the accuracy and reliability of pulmonary disease detection from chest X-ray images. The ability to distinguish between conditions such as COVID-19, pneumonia, and tuberculosis with high precision and recall can be pivotal in clinical settings, leading to faster and more accurate diagnoses, ultimately improving patient care outcomes.

The promising results achieved in chest X-ray classification through our use of MobileNet architecture, complemented by Energy Valley Optimization (EVO), signal a broader applicability to a variety of medical image analysis tasks. The ability of MobileNet to efficiently process and accurately classify images holds great potential beyond the realm of pulmonary diseases. For instance, its application could be extended to the analysis of histopathological slides in oncology, where the differentiation between benign and

malignant cellular structures is crucial. Similarly, the adaptability of this framework could be beneficial in the examination of MRI or CT scans for the detection of neurological anomalies, such as strokes or tumors, where the detailed and complex nature of the imagery requires both nuanced feature extraction and rapid processing.

The core principles of our approach, namely the streamlined efficiency of MobileNet and the precise hyperparameter tuning enabled by EVO, offer a foundation that can be tailored to various imaging modalities and diagnostic requirements. By adjusting the convolutional filters and training procedures, the framework can be optimized for the specific characteristics of different image types, whether they require the capture of minute textural differences in dermatological imaging or the delineation of intricate vascular structures in angiography. Furthermore, the capacity of MobileNet to function effectively even with limited computational resources makes it a versatile tool for medical facilities with varying levels of technological infrastructure. This adaptability, coupled with the model's high accuracy, provides a valuable avenue for enhancing diagnostic capabilities and ultimately, patient care across multiple domains of medical imaging.

The nuanced precision-recall balance elucidated in our analysis is particularly critical for medical diagnostics, where the ability to correctly identify conditions with shared radiographic features is paramount. For instance, the model's high recall for classes such as 'mass' (0.93) and 'nodule' (0.83), albeit at the expense of lower precision, is of considerable clinical importance. The imperative to minimize false negatives in these categories, where a missed diagnosis could have grave implications, justifies the acceptance of more false positives, which, while not ideal, are less detrimental in a clinical context given the possibility of follow-up testing.

F1-scores, as seen in our data ranging from 0.79 for 'Effusion' to 0.95 for 'Tuberculosis' post-EVO in MobileNet, along with AUC values close to 1.00 across most classes in the ROC curves, provide a robust measure of the model's diagnostic accuracy. These metrics are integral to evaluating the model's efficacy, with the F1-score offering a balanced view of precision and recall, and the AUC reflecting the model's discriminative power.

The uniformity in sensitivity and specificity, notably achieving a perfect 1.00 in some classes, while admirable, raises the specter of overfitting or data imbalance. To counteract this, our future work will involve more rigorous validation techniques and possibly the introduction of additional data sources to ensure the model's reliability across diverse clinical scenarios. The model's balanced performance across various pathologies is not only a testament to its robustness but also to its potential for equitable application across different diseases, preventing the overshadowing of less common diseases by more prevalent ones. This is essential for ensuring comprehensive patient care.

Post-EVO, the uplift in MobileNet's performance is evident, with precision, recall, and F1-scores showing significant im- provements, such as the precision for 'Lung Opacity' increas- ing from 0.95 to 1.00, and the F1-score for 'Pneumothorax' improving from 0.94 to 0.91. Such enhancements are indicative of the fine-tuning that EVO provides, optimizing the model to deliver more precise and clinically actionable results.

In sum, the post-EVO model presents as a highly capable diagnostic tool, with precision values up to 0.99 for 'COVID' and 'Mass', recall rates peaking at 1.00 for 'Control' and 'COVID', and an overall F1-score of 0.94, suggesting a finely- tuned balance between sensitivity and specificity. The future trajectory of our research will focus on further refining these metrics, broadening the model's diagnostic scope, and ensuring its application is as efficacious in real-world settings as it is within the controlled conditions of our study.

VIII. CONCLUSION

Our investigation into the application of deep learning for the classification of pulmonary diseases through chest X-ray imagery has yielded insightful and encouraging results. We compared the efficacy of a standard Convolutional Neural Network (CNN) against the MobileNet architecture, both prior to and following the application of Energy Valley Optimization (EVO) for hyperparameter tuning.

The CNN provided a strong baseline with an overall accuracy of 85.91%, which was markedly enhanced to 93.30% by the MobileNet architecture even before optimization. The incorporation of EVO further augmented MobileNet's performance, culminating in a post-optimization accuracy of 94.18%. These improvements were also mirrored in the precision, recall, and F1-score metrics across the various classes.

The application of deep learning, particularly the refined MobileNet model post-EVO, has demonstrated potential as a robust tool for medical diagnostics, achieving near-perfect sensitivity and specificity. This suggests that such models can be reliable adjuncts to traditional diagnostic methods, potentially expediting the diagnostic process and improving the accuracy of clinical outcomes.

Furthermore, the high AUC scores obtained from our ROC analysis underscore the capability of the optimized MobileNet architecture to accurately classify a wide range of pulmonary conditions. This holds promise for the broader domain of medical image analysis, where such models can be adapted for different types of imaging and disease detection tasks.

In summary, the implications of our findings are substantial for the field of automated medical image analysis. They indicate a step forward in the development of highly accurate, efficient, and generalizable diagnostic tools that can support and enhance clinical decision-making. Future directions for this research include expanding the dataset size, addressing potential class imbalances, and incorporating extensive clinical validation. This will help to ensure that the models not only perform well computationally but also align closely with real-world clinical diagnostic processes.

REFERENCES

[1] S. Albahli, "Efficient gan-based chest radiographs (cxr) augmentation to diagnose coronavirus disease pneumonia," *International Journal of Medical Sciences*, vol. 17, pp. 1439–1448, 2020.

[2] K. H. Shibly, S. K. Dey, M. T. U. Islam, and M. M. Rahman, "Covid faster r-cnn: A novel framework to diagnose novel coronavirus disease (covid-19) in x-ray images," *Informatics in Medicine Unlocked*, vol. 20, p. 100405, 2020.

[3] S. Basu and R. H. Campbell, "Going by the numbers: Learning and modeling covid-19 disease dynamics," *Chaos, Solitons & Fractals*, vol. 138, p. 110140, 2020.

[4] P. K. Sethy and S. K. Behera. *Detection of coronavirus disease (covid-19) based on deep features*. Preprints, 2020.

[5] S. Hassantabar, M. Ahmadi, and A. Sharifi, "Diagnosis and detection of infected tissue of covid-19 patients based on lung x-ray image using convolutional neural network approaches," *Chaos, Solitons & Fractals*, vol. 140, p. 110170, 2020.

[6] D. A. Ragab, M. Sharkas, S. Marshall, and J. Ren, "Breast cancer detection using deep convolutional neural networks and support vector machines," *PeerJ*, vol. 7, p. e6201, 2019.

[7] Z. Yao, J. Li, Z. Guan, Y. Ye, and Y. Chen, "Liver disease screening based on densely connected deep neural networks," *Neural Networks*, vol. 123, pp. 299– 304, 2020.

[8] I. Pacal, D. Karaboga, A. Basturk, B. Akay, and U. Nalbantoglu, "A comprehensive review of deep learning in colon cancer," *Computers in Biology and Medicine*, vol. 126, p. 104003, 2020.

[9] X. W. Gao, R. Hui, and Z. Tian, "Classification of ct brain images based on deep learning networks," *Computer Methods and Programs in Biomedicine*, vol. 138, pp. 49–56, 2017.

[10] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017.

[11] W. Ausawalaithong, A. Thirach, S. Marukatat, and T. Wilaiprasitporn, "Automatic lung cancer prediction from chest x-ray images using the deep learning approach," in *2018 11th Biomedical Engineering International Conference (BMEiCON)*, pp. 1–5, 2018.

[12] N. M. Elshennawy and D. M. Ibrahim, "Deep-pneumonia framework using deep learning models based on chest x-ray images," *Diagnostics*, vol. 10, p. 649, 2020.

[13] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of covid- 19 in chest x-ray images using detrac deep convolutional neural network," *arXiv preprint arXiv:2003.13815*, 2020.

[14] S. Asif, Y. Wenhui, H. Jin, and S. Jinhai, "Classification of COVID-19 from Chest X-ray images using Deep Convolutional Neural Network," *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, pp. 426-433, 2020, doi: 10.1109/ICCC51575.2020.9344870.

[15] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images," *arXiv preprint arXiv:2003.11055*, 2020.

[16] C. Zheng *et al*., "Deep learning-based detection for covid-19 from chest ct using weak label," *medRxiv*, 2020.

[17] J. Zhao, Y. Zhang, X. He, and P. Xie, "Covid-ct-dataset: A ct scan dataset about covid-19," *arXiv preprint arXiv:2003.13865*, 2020.

[18] O. Gozes, M. Frid-Adar, N. Sagie, H. Zhang, W. Ji, and H. Greenspan, "Coronavirus detection and analysis on chest ct with deep learning," *arXiv preprint arXiv:2004.02640*, 2020.

[19] S. Wang *et al*., "A deep learning algorithm using ct images to screen for coronavirus disease (covid-19)," *European radiology*, vol. 31, pp. 6096-6104, 2020.

[20] L. Li *et al*., "Artificial intelligence distinguishes covid-19 from community- acquired pneumonia on chest ct," *Radiology*, vol. 296, no. 2, pp. 65–71, 2020.

[21] J. Chen *et al*., "Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography," *Scientific Reports*, vol. 10, pp. 1–11, 2020.

[22] B. Wang *et al*., "Ai-assisted ct imaging analysis for covid-19 screening: building and deploying a medical ai system," *Applied Soft Computing*, vol. 98, p. 106897, 2020.

[23] M. Yamac, M. Ahishali, A. Degerli, S. Kiranyaz, M. E. Chowdhury, and M. Gabbouj, "Convolutional sparse support estimator based covid-19 recognition from x-ray images," *arXiv preprint arXiv:2005.04014*, 2020.

[24] Y. Oh, S. Park, and J. C. Ye, "Deep Learning COVID-19 Features on CXR Using Limited Training Data Sets," in *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2688-2700, Aug. 2020, doi: 10.1109/TMI.2020.2993291.

[25] M. Loey, F. Smarandache, and N. E. M. Khalifa, "Within the lack of chest covid-19 x-ray dataset: a novel detection model based on gan and deep transfer learning," *Symmetry*, vol. 12, p. 651, 2020.

[26] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of covid-19 cases using deep neural networks with x-ray images," *Computers in biology and medicine*, vol. 121, p. 103792, 2020.

[27] N. E. M. Khalifa, M. H. N. Taha, A. E. Hassanien, and S. Elghamrawy, "Detection of coronavirus (covid-19) associated pneumonia based on generative adversarial networks and a fine-tuned deep transfer learning model using chest x-ray dataset," in *International Conference on Advanced Intelligent Systems and Informatics*, pp. 234-247, 2022.

[28] A. A. Ardakani, A. R. Kanafi, U. R. Acharya, N. Khadem, and A. Mohammadi, "Application of deep learning technique to manage covid-19 in routine clinical practice using ct images: results of 10 convolutional neural networks," *Computers in biology and medicine*, vol. 121, p. 103795, 2020.

[29] H. S. Maghdid, A. T. Asaad, K. Z. Ghafoor, A. S. Sadiq, and M. K. Khan, "Diagnosing covid-19 pneumonia from x-ray and ct images using deep learning and transfer learning algorithms," *arXiv preprint arXiv:2004.00038*, 2020.

[30] T. Yan, P.K. Wong, H. Ren, H. Wang, J. Wang, and Y. Li, "Automatic distinction between covid-19 and common pneumonia using multi-scale convolutional neural network on chest ct scans," *Chaos, Solitons Fractals*, vol. 140, p. 110153, 2020.

[31] V. Perumal, V. Narayanan, and S. J. S. Rajasekar, "Detection of covid-19 using cxr and ct images using transfer learning and haralick features," *Applied Intelligence*, vol. 51, pp. 341-358, 2020.

[32] A. Guarnera, E. Santini, and P. Podda, "Covid-19 pneumonia and lung cancer: A challenge for the radiologist review of the main radiological features, differential diagnosis and overlapping pathologies," *Tomography*, vol. 8, no. 1, pp. 513–528, 2022.

[33] R. Li, C. Xiao, Y. Huang, H. Hassan, and B. Huang, "Deep learning applications in computed tomography images for pulmonary nodule detection and diagnosis: A review," *Diagnostics*, vol. 12, no. 2, p. 298, 2022.

[34] H. Farhat, G. E. Sakr, and R. Kilany, "Deep learning applications in pulmonary medical imaging: recent updates and insights on covid-19," *Machine vision and applications*, vol. 31, pp. 1–42, 2020.

[35] S. Bharati, P. Podder, M. Mondal, and V. B. Prasath, "Medical imaging with deep learning for covid-19 diagnosis: a comprehensive review," *arXiv preprint arXiv:2107.09602*, 2021.

[36] B. Liu, W. Chi, X. Li, P. Li, W. Liang, H. Liu, and J. He, "Evolving the pulmonary nodules diagnosis from classical approaches to deep learning-aided decision support: three decades' development course and future prospect," *Journal of cancer research and clinical oncology*, vol. 146, pp. 153–185, 2020.

[37] F. Bushra *et al*., "Deep learning in computed tomography pulmonary angiography imaging: A dual-pronged approach for pulmonary embolism detection," *Expert Systems with Applications*, vol. 245, p. 123029, 2024.

[38] M. A. Talukder, M. A. Layek, M. Kazi, M. A. Uddin, and S. Aryal, "Empowering covid-19 detection: Optimizing performance through fine-tuned efficientnet deep learning architecture," *Computers in Biology and Medicine*, vol. 168, p. 107789, 2024.

[39] M. V. Sanida, T. Sanida, A. Sideris, and M. Dasygenis, "An advanced deep learning framework for multi-class diagnosis from chest x-ray images," *J*, vol. 7, no. 1, pp. 48–71, 2024.

[40] S. Umirzakova, S. Ahmad, L. U. Khan, and T. Whangbo, "Medical image super resolution for smart healthcare applications: A comprehensive survey," *Information Fusion*, p. 102075, 2023.

[41] S. Atasever, N. Azginoglu, D. S. Terzi, and R. Terzi, "A comprehensive survey of deep learning research on medical image analysis with focus on transfer learning," *Clinical Imaging*, vol. 94, pp. 18–41, 2023.

[42] T. Mahmood, A. Rehman, T. Saba, L. Nadeem, and S. A. O. Bahaj, "Recent Advancements and Future Prospects in Active Deep Learning for Medical Image Segmentation and Classification," in *IEEE Access*, vol. 11, pp. 113623-113652, 2023, doi: 10.1109/ACCESS.2023.3313977.

[43] M. Lubbad, D. Karaboga, A. Basturk, B. Akay, U. Nalbantoglu, and I. Pacal, "Machine learning applications in detection and diagnosis of urology cancers: a systematic literature review," *Neural Computing and Applications*, pp. 1–25, 2024.

[44] N. Naik *et al.*, "Role of deep learning in prostate cancer management: past, present and future based on a comprehensive literature Review," *Journal of Clinical Medicine*, vol. 11, no. 13, p. 3575, 2022.

[45] H. E. Sarog˘lu *et al*., "Machine learning, iot and 5g technologies for breast cancer studies: A review," *Alexandria Engineering Journal*, vol. 89, pp. 210– 223, 2024.

[46] D. Tian *et al*., "A review of traditional chinese medicine diagnosis using machine learning: Inspection, auscultation-olfaction, inquiry, and palpation," *Computers in Biology and Medicine*, p. 108074, 2024.

[47] S. Wassan *et al*, "Deep convolutional neural network and iot technology for healthcare," *Digital Health*, vol. 10, 2024.

[48] K. Sarkar, A. Shiuly, and K. G. Dhal, "Revolutionizing concrete analysis: An in-depth survey of ai-powered insights with image-centric approaches on comprehensive quality control, advanced crack detection and concrete property exploration," *Construction and Building Materials*, vol. 411, p. 134212, 2024.

[49] S. K. Zhou *et al.*, "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 820– 838, 2021.

[50] X. Chen, "Recent advances and clinical applications of deep learning in medical image analysis," *Medical Image Analysis*, vol. 79, p. 102444, 2022.

[51] A. S. Panayides *et al*., "Ai in medical imaging informatics: current challenges and future directions," *IEEE journal of biomedical and health informatics*, vol. 24, no. 7, pp. 1837–1857, 2020.

[52] D. Karimi, H. Dou, S. K. Warfield, and A. Gholipour, "Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis," *Medical Image Analysis*, vol. 65, p. 101759, 2020.

[53] A. Singh, S. Sengupta, and V. Lakshminarayanan, "Explainable deep learning models in medical image analysis," *Journal of imaging*, vol. 6, no. 6, p. 52, 2020.

[54] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, and X. Ding, "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation," *Medical Image Analysis*, vol. 63, p. 101693, 2020.

[55] M. B. Sushma and S. Prusty, "The effect of covid-19 on public transportation sectors and conceptualizing the shifting paradigm: A report on indian scenario," in *Advances in Sustainable Materials and Resilient Infrastructure*, pp. 21–40, 2022.

[56] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks," *arXiv preprint arXiv:2003.10849*, 2020.

[57] M. E. Chowdhury *et al*., "Can ai help in screening viral and covid-19 pneumonia?," *arXiv preprint arXiv:2003.13145*, 2020.

[58] X. Xu *et al*., "A deep learning system to screen novel coronavirus disease 2019 pneumonia," *Engineering*, vol. 6, no. 10, pp. 1122–1129, 2020.

[59] R. M. Pereira, D. Bertolini, L. O. Teixeira, C. N. Silla Jr., and Y. M. Costa, "Covid-19 identification in chest x-ray images on flat and hierarchical classification scenarios," *Computational Methods and Programs in Biomedicine*, p. 105532, 2020.

[60] J. Zhang, Y. Xie, Y. Li, C. Shen, and Y. Xia, "Covid-19 screening on chest x-ray images using deep learning based anomaly detection," *arXiv preprint arXiv:2003.12338*, 2020.