[1] K. Kalaiselvi
[2] Dr. M. Kasthuri

Pneumonia Classification using VGG19 Transfer Learning and Data Augmentation with PCA Across Diverse Chest X-Ray Datasets

JES Journal of Electrical Systems

Abstract: **-** Pneumonia remains a significant global health issue, particularly in low-resource settings, making the development of efficient and accurate diagnostic tools crucial. This study presents a deep learning-based approach using transfer learning with a fine-tuned VGG19 model for pneumonia classification from chest X-ray images. The model was tested across three different datasets (D1, D2, D3) to evaluate its performance in both within-dataset and cross-dataset scenarios. Preprocessing involved resizing the images, followed by data augmentation to improve model generalization. Principal Component Analysis (PCA) was applied to reduce the dimensionality of features extracted from the convolutional layers, optimizing both training time and accuracy. The model achieved high classification accuracies, with performance exceeding **99%** in scenarios where the training and testing datasets were identical, as demonstrated by results on D1 and D2. However, cross-dataset evaluation, particularly when training on D1 and testing on D3, revealed a noticeable drop in accuracy, with a minimum of **79.25%** in some metrics. The training time ranged from **00:00:15** to **00:13:08**, depending on the dataset complexity, demonstrating that PCA.

Keywords: CXR images, Transudative Transfer Learning, VGG19 model, PCA, SoftMax.

I. INTRODUCTION

Pneumonia remains one of the leading causes of global mortality, disproportionately affecting vulnerable populations such as children and the elderly. Traditional diagnostic methods, which often depend on the interpretation of chest X-rays (CXR) by experienced radiologists, can be subjective, time-consuming, and prone to variability. In recent years, deep learning techniques, especially Convolutional Neural Networks (CNNs), have emerged as powerful tools for automating medical image classification, with significant potential in diagnosing pneumonia.

The VGG19 CNN model, recognized for its outstanding performance in general image classification tasks, has been extensively applied across various domains. However, when applied to medical imaging, its effectiveness can be delayed by the limited availability of labeled datasets in healthcare. To overcome this limitation, inductive transfer learning (ITL) enables pre-trained models like VGG19 to be adapted to specific medical tasks, improving performance while maintaining the models generalization capabilities.

This study seeks to classify pneumonia using CXRs from three distinct datasets, leveraging a pre-trained VGG19 model fine-tuned with ITL. Additionally, PCA and a SoftMax classifier are integrated to further optimize the classification process. By refining the VGG19 model with these methods, this study aims to enhance the accuracy of automated pneumonia diagnosis across multiple CXR datasets, potentially contributing to more reliable clinical decision-making and reducing diagnostic errors.

II. REVIEW OF LITERATURE

Pneumonia diagnosis using CXR images has been a focal point of Deep Learning (DL) research due to the global burden of the disease. Traditional diagnostic approaches rely heavily on the expertise of radiologists, but DL methods have demonstrated the potential to improve diagnostic accuracy and efficiency by automating the detection process. CNNs are the backbone of these methods, offering the capability to extract spatial hierarchies in images, which are crucial for medical imaging tasks. Studies such as Rajpurkar et al. (2017) have used CNNs like CheXNet to detect pneumonia from CXR images with promising results, showing near-expert level performance.

Transfer learning (TL) has emerged as a powerful tool for overcoming the challenges posed by limited labeled data in the medical domain. By leveraging pre-trained models like VGG19, originally trained on large general-purpose datasets like ImageNet, TL allows fine-tuning of these models to perform specialized tasks such as pneumonia detection from CXRs. Research by Shin et al. (2016) highlights the advantages of TL in medical

[1] Research Scholar, Bishop Heber College, (Affiliated to Bharathidasan University), Tiruchirappalli, Tamil Nadu, India.
kalaiselvigl1984@gmail.com
[2] Associate Professor, Bishop Heber College, (Affiliated to Bharathidasan University), Tiruchirappalli, Tamil Nadu, India.
kasthuri.ca@bhc.edu.in

imaging, particularly when data scarcity is a major hurdle. Transudative transfer learning, a specific type of TL, further enhances model generalizability by adapting the model more effectively to the target domain, making it well-suited for tasks where labeled data is scarce, as demonstrated by Bai et al. (2020) in medical image segmentation.

The VGG19 architecture, developed by Simonyan and Zisserman (2014), is widely recognized for its deep and efficient feature extraction capabilities, making it suitable for complex image classification tasks. Its success in ImageNet classification has encouraged its use in medical imaging applications, including CXR analysis. Zhang et al. (2020) successfully employed VGG19 for pneumonia detection from CXRs, achieving high classification accuracy by fine-tuning the pre-trained model on medical datasets. Despite its computational intensity, VGG19's depth makes it ideal for capturing intricate features in medical images.

Deep learning models like VGG19 extract a vast number of features from images, often resulting in high-dimensional feature spaces. To mitigate the computational burden and reduce overfitting, dimensionality reduction techniques like PCA are commonly used. PCA helps retain the most relevant information by transforming the original features into a smaller set of uncorrelated components, as explored by Abdi and Williams (2010). In medical image analysis, PCA has been effectively used to simplify the feature space without compromising classification accuracy, enabling faster and more efficient model training.

The SoftMax classifier, a widely used function in multi-class classification tasks, has become a standard component of DL pipelines. It maps the model's output into a probability distribution, ensuring that the sum of the output probabilities equals one, making it suitable for classification tasks like pneumonia detection. The success of CNNs in medical image classification often hinges on the final layer's performance, where the SoftMax classifier determines the class label. Several studies, including the work of Esteva et al. (2017) on dermatological conditions, have demonstrated the efficacy of combining CNN feature extraction with SoftMax classifiers for high-accuracy medical diagnoses.

Training and validating models across multiple datasets is essential for ensuring that the model generalizes well to various data distributions. In the field of medical imaging, studies have employed multi-dataset approaches to improve the robustness of diagnostic models. For example, Irvin et al. (2019) used a multi-institutional dataset for CXR analysis to better generalize their pneumonia detection model across different clinical settings. Leveraging three distinct datasets in this study not only helps in assessing the adaptability of the VGG19 model but also enhances the model's performance across diverse clinical environments.

III. METHODOLOGY

The methodology for pneumonia classification involves three datasets of chest X-ray images, preprocessing, data augmentation, a pretrained VGG19 model, fine-tuning with PCA, and performance evaluation. The images are resized to 224x224 pixels, and the model is fine-tuned using VGG19 and PCA to further reduce feature dimensionality. The model's effectiveness in pneumonia classification is assessed using metrics like accuracy and precision. Figure shows the framework of PICILTFA-PCA: SoftMax model.

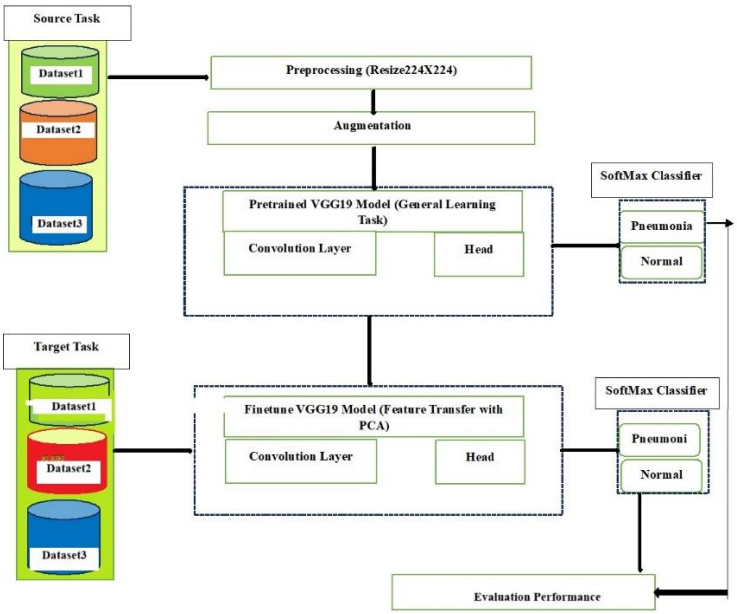

Figure1: PICILTFA-PCA: SoftMax model.

3.1. Datasets

This study uses three distinct datasets for pneumonia classification, referred to as D1, D2, and D3. These datasets consist of chest X-ray (CXR) images, split into training and testing sets with a 60%-40% ratio.

Dataset D1(from Kaggle) contains a total of 5,216 images, including 1,341 normal images and 3,875 pneumonia images. The training set includes 2,325 pneumonia and 805 normal images, while the testing set includes 1,550 pneumonia and 536 normal images [10].

Dataset D2(from Roboflow) contains 4,077 images, including 1,104 normal and 2,973 pneumonia images. The training set consists of 1,784 pneumonia and 663 normal images, with the testing set having 1,189 pneumonia and 441 normal images [11].

Dataset D3(from Kaggle) contains 177 images, with 106 normal and 71 pneumonia images. The training set comprises 42 pneumonia and 64 normal images, while the testing set has 28 pneumonia and 43 normal images [12].

3.2. Preprocessing and Data Augmentation

Before model training, the CXR images are preprocesses to ensure uniformity. The following steps are performed:

Resizing: All images are resized to 224x224 pixels to match the input requirements of the VGG19 model.

Normalization: Pixel values are scaled to the range [0, 1] to standardize input data.

Grayscale Conversion: Since CXRs are grayscale images, they are converted to a 3-channel format required by the VGG19 architecture.

Data Augmentation: To artificially expand the training data and improve model generalization, augmentation techniques are applied. These include random rotations (±30 degrees), horizontal flipping, zooming (80%-120%), and brightness adjustment (90%-110%).

3.3. Model Architecture: CustomisedVGG19

The VGG19 architecture, pre-trained on ImageNet, is selected due to its strong feature extraction capabilities. The model consists of 19 layers, including convolutional layers, max-pooling layers, and fully connected layers. Given the complexity of pneumonia detection in medical images, VGG19 is fine-tuned for the specific task of pneumonia classification across the three datasets.

Transfer Learning: The initial layers of VGG19, responsible for low- and mid-level feature extraction, are frozen to retain the knowledge from Pneumonia CXR dataset. Only the deeper layers are fine-tuned to adapt to the medical domain, leveraging the pre-trained features while adjusting to the specific characteristics of the CXR datasets.

3.4. Inductive Transfer Learning

ITL is employed to transfer knowledge from the source domain (Pneumonia CXR dataset) to the target domains (D1, D2, D3). Unlike traditional transfer learning, Inductive learning aims to improve model generalization across same datasets by leveraging unlabelled data or limited labeled samples from the target domain. This method ensures that the model is not overfitted to a single dataset but instead learns to generalize across the three datasets.

3.5. Dimensionality Reduction with Principal Component Analysis (PCA)

After extracting features from the fine-tuned VGG19 model, PCA is applied to reduce the dimensionality of the feature set. PCA transforms the high-dimensional feature space into a lower-dimensional space by selecting the principal components (100) that explain the most variance in the data. This step reduces the computational complexity of the classification task while retaining the most relevant information for pneumonia detection.

Feature Extraction: The output of the last convolutional layer of VGG19 is flattened to a feature vector.

PCA: Dimensionality is reduced based on a predefined threshold, retaining around 95% of the data variance while reducing noise and redundancy in the feature set.

3.6. Classification: SoftMax Classifier

A SoftMax classifier is employed as the final classification layer to categorize the CXR images into "pneumonia" or "normal." The reduced feature vectors from the PCA stage are fed into the SoftMax layer, which outputs the probability distribution across the two classes.

Optimization: The Adam optimizer is used to minimize the cross-entropy loss function, and model training is performed using backpropagation.

3.7. Model Training and Validation

The model is trained and validated using the three datasets:

Training-Validation Split: Each dataset is split into 60% training and 40% validation sets. Cross-dataset validation is also conducted by training on one dataset and validating on the others to assess generalizability.

Training Parameters: The model is trained for 10 epochs with a batch size of 128, using a learning rate of 0.0001[9].

Early Stopping: Early stopping is employed to prevent overfitting by monitoring the validation loss.

3.8. Evaluation Metrics

The model's performance is evaluated using several metrics.

Accuracy: The percentage of correctly classified images.

Precision: The proportion of true positive pneumonia detections out of all positive predictions.

Recall (Sensitivity): The proportion of true positive pneumonia cases identified out of all actual pneumonia cases.

Specificity: The proportion of true negative pneumonia cases identified out of all actual pneumonia cases

F1 Score: The harmonic means of precision and recall.

IV. RESULT

The results show that the model's performance varies based on the dataset combinations. The model performs best when trained and tested on the same dataset (e.g., **CVGGPCILAPCAD1D1**, **CVGGPCILAPCAD2D2**), with accuracies consistently above **99%**, reflecting strong feature learning and classification capabilities. On the other hand, cross-dataset evaluation, particularly with D3, reveals that the model struggles more with generalization, especially when trained on D1 and tested on D3 (e.g., **CVGGPCILAPCAD1D3**), with performance dropping significantly. The application of **PCA** for feature extraction and dimensionality reduction, alongside **VGG19** transfer learning, demonstrates a clear benefit in reducing computational time, as evidenced by the short training times for D3, while maintaining relatively high classification accuracy. However, the variability in performance when applying the model to different datasets suggests that additional fine-tuning or data augmentation may be required to improve generalization, particularly when testing on unseen or distinct datasets.

Table1: CVGGPCILPCA: SoftMax Model

Algorithm	Dataset	Training Time (hh:mm: ss)	CVGGPCILAPCAV1	CVGGPCILPCAV2	CVGGPCILAPCAV3	CVGGPCILAPCAV4	CVGGPCILAPCAV5
CVGGPCILAD1D1	D1, D1	00:12:45	0.9895	0.9646	0.9981	0.9879	0.9895
CVGGPCILAPCAD2D1	D2, D1	00:13:08	0.9963	0.9955	0.9966	0.9983	0.9975
CVGGPCILAPCAD3D1	D3, D1	00:00:21	0.9083	0.7925	1.0000	0.8590	0.9241
CVGGPCILAPCAD1D2	D1, D2	00:09:08	0.9866	0.9590	0.9961	0.9860	0.9910
CVGGPCILAPCAD2D2	D2, D2	00:12:32	0.9939	0.9977	0.9924	0.9992	0.9958
CVGGPCILAPCAD3D2	D3, D2	00:00:22	0.9500	0.9623	0.9403	0.9692	0.9545
CVGGPCILAPCAD1D3	D1, D3	00:07:28	0.9751	0.9049	0.9958	0.9681	0.9835
CVGGPCILAPCAD2D3	D2, D3	00:11:52	0.9721	0.9005	0.9966	0.9642	0.9801
PCILAPCAD3D3	D3, D3	00:00:15	0.9583	0.9434	0.9701	0.9559	0.9630

Training progress CVGGPICILTFA-PCA: SoftMax model

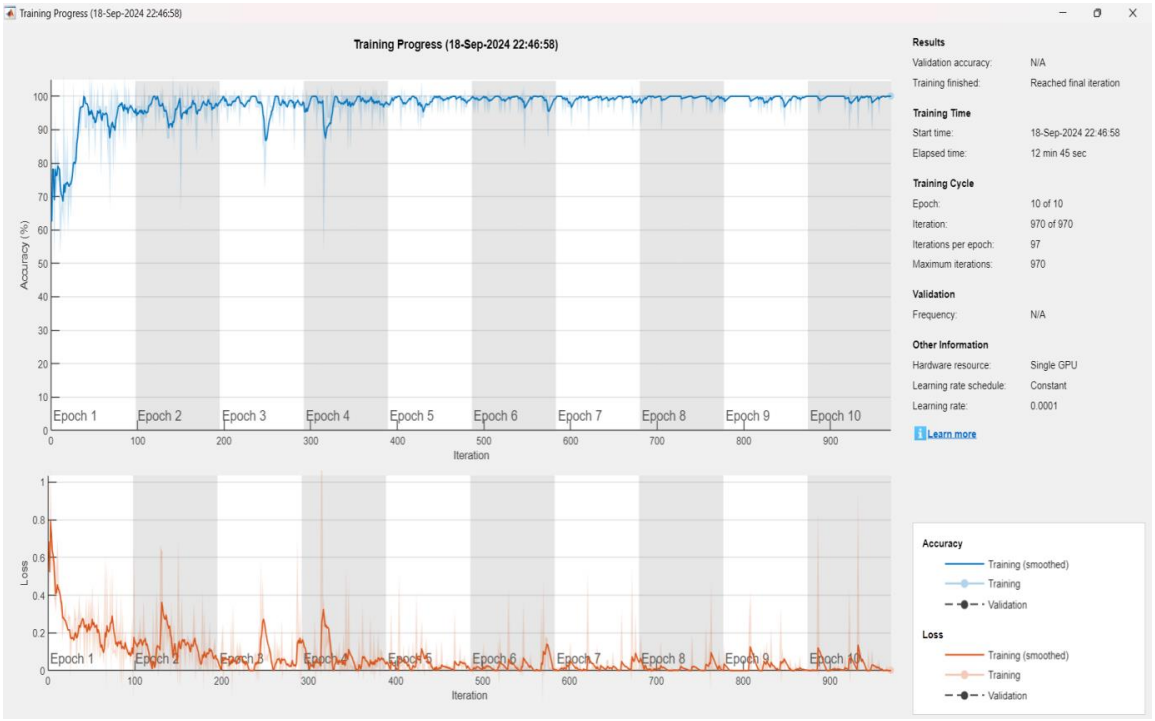

Figure2: PICILTFA-PCA:SoftMax model

Confusion Matrix

CVGGPCILAPCA			
TARGET / OUTPUT	Pneumonia	Normal	SUM
Pneumonia	1547 74.16%	19 0.91%	1566 98.79% 1.21%
Normal	3 0.14%	517 24.78%	520 99.42% 0.58%
SUM	1550 99.81% 0.19%	536 96.46% 3.54%	2064 / 2086 98.95% 1.05%

V. CONCLUSION

The use of a fine-tuned VGG19 model with PCA shows strong potential for pneumonia classification, particularly when training and testing on the same dataset. However, cross-dataset variability suggests the need for further refinement to improve generalization across diverse datasets. The overall approach achieves high accuracy while keeping training times manageable, especially with PCA's feature reduction, making this model suitable for practical applications in pneumonia detection from chest X-rays.

REFERENCES

[1] Rajpurkar, P., et al. (2017). "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning." arXiv preprint arXiv:1711.05225.

[2] Shin, H.-C., et al. (2016). "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics, and Transfer Learning." IEEE Transactions on Medical Imaging, 35(5), 1285-1298.

[3] Bai, W., et al. (2020). "Self-Supervised Learning for Cardiac MR Image Segmentation by Anatomical Position Prediction." Medical Image Analysis, 63, 101703.
[4] Zhang, J., et al. (2020). "Pneumonia Detection from Chest X-Ray Images Based on Convolutional Neural Networks." BioMed Research International, 2020, 647536.
[5] Abdi, H., & Williams, L.J. (2010). "Principal Component Analysis." Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), 433-459.
[6] Lakshmi, S., & Martin, D. (2019). "PCA Based Dimensionality Reduction in Medical Image Analysis." Journal of Medical Systems, 43, 282.
[7] Esteva, A., et al. (2017). "Dermatologist-level classification of skin cancer with deep neural networks." Nature, 542(7639), 115-118.
[8] Irvin, J., et al. (2019). "CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison." Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), 590-597.
[9] Kalaiselvi, K., & Kasthuri, M. (2024). Tuning VGG19 hyperparameters for improved pneumonia classification. *The Scientific Temper*, *15*(02), 2231–2237.
[10] https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia.
[11] https://universe.roboflow.com/mohamed-traore-2ekkp/chest-x-rays-qjmia/dataset/2.
[12] https://www.kaggle.com/datasets/marcoferoldi/chest-xray-small?resource=download.