https://doi.org/10.31449/inf.v49i4.7751 Informatica 49 (2025) 825–840 825

Comparative Analysis of Transfer Learning and Few-Shot Learning with CNN Architectures for Chest X-Ray Classification under Data Constraints

Sourav Paul[1,2], Ranjita Das[3], Vaibhav Malviya[2], Anurag Mhatre[1]
[1]Department of Computer Science & Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majhitar, India
[2]Department of Computer Science & Engineering, National Institute of Technology Mizoram, Aizawl, India
[3]Department of Computer Science & Engineering, National Institute of Technology Agartala, Agartala, India
E-mail: sourav.cst@gmail.com, ranjita.nitm@gmail.com, vaibhav.cse@nitmz.ac.in, anuragmhatre.2002@gmail.com

Keywords: Deep learning, transfer learning, image classification, X-ray

Received: December 5, 2024

This study focuses on the early and accurate diagnosis of life-threatening lung diseases such as COVID-19, pneumonia, and lung opacity using deep learning. Since deep learning requires large datasets that are often limited in medical imaging, the work applies transfer learning to overcome this challenge. Six pre-trained CNN models—VGG19, VGG16, ResNet50, MobileNetV2, InceptionV3, and DenseNet201—are used to classify chest X-ray images through feature extraction and fine-tuning techniques. In the evaluation phase, a range of classifiers, including Random Forest, K-Nearest Neighbors, Extra Trees, and Decision Tree, were employed to assess the predictive capabilities of the CNN-derived features. The outcomes reveal insights into the compatibility of these classifiers with different transfer learning strategies. Furthermore, this study delves into the realm of few shot learning, utilizing a limited subset of 15 images from each class. The efficacy of both transfer learning and few-shot learning in the context of this constrained dataset is examined, shedding light on the adaptability of these techniques to scenarios with limited training samples. The results showcase the strengths and limitations of each approach, providing valuable insights into the intricate landscape of chest X-ray classification.Results show that for the dataset having a total of 3707 images comprising four different classes, the fine-tuned method has outperformed the feature-extracted method for all the deep learning models executed, giving a high accuracy of 98.89% for the DenseNet201 model with data augmentation and Extra Tree classifier. For the case where only 15 images have been taken from each of the four classes, Siamese Networks type few-shot learning has outperformed both a base model and two types of transfer learning models, yielding the best accuracy of 96.84% for the DenseNet 201 model.This work contributes to the ongoing efforts to develop reliable and efficient diagnostic tools amidst the evolving challenges posed by the recent COVID-19 pandemic.

Povzetek: Primerjano je prenosno učenje in učenje iz malo primerov za klasifikacijo rentgenskih posnetkov prsnega koša (COVID-19, pljučnica, pljučna motnost, normalno) pri pomanjkanju podatkov. Pri 3707 slikah najboljše rezultate doseže prilagojeni DenseNet201 z augmentacijo in Extra Trees, pri 15 slikah na razred pa Siamese few-shot pristop.

1 Introduction

According to a study[1] that examined the global burden of lung illness , it is the third leading cause of mortality worldwide. One of the simple, cost-effective, and noninvasive modes of testing for lung diseases is chest X-ray [16]. But the unavailability of trained radiologists triggers the use of computer vision to identify lung diseases from chest X-rays.Convolutional neural networks are now the most advanced method for resolving image classification issues in a range of industries, including biology [5], security[17] , and medicine[12][3] .It may be difficult to acquire and annotate the thousands of images that deep learning algorithms demand[4] .At the present time the CNN algorithms that are being used to classify the images use a lot of training parameters; as a result, they require a lot of training data as well as huge hardware requirements, which are not available most of the time[15].To address these problems, in this paper we tested the performances of transfer learning, data augmentation, and a few-shot learning techniques. The performance of ML-based classifiers, which take very few parameters as compared to deep neural networks, has also been experimented with.This paper embarks on a multifaceted exploration of AI-driven chest X-ray classification, encompassing CNN models, transfer learning strategies, ensemble classifiers, and few-shot learning. The ultimate aim is to unravel the intricacies of applying cutting-edge techniques to medical image analysis in the context of lung disease diagnosis. Through the synthesis of empirical findings and insightful interpretations, this research contributes to the ongoing endeavor to develop accurate, efficient, and scalable diagnostictools in the face of the global COVID-19 pandemic.The main contributions of this work are as follows.

1. How to run a deep learning model on insufficient medical images for classification.

2. Experimented with the variation of Transfer learning.

3. Tested the application of Few-shot learning on a very short data set.

4. Checked the compatibility of different types of ML-based classifiers with Deep learning models.

2 Related work

Author[7] has classified thorax xray to COVID-19, Pneumonia, and Healthy cases applying VGG19 with transfer learning.first 16 layer of Pretrained VGG19 has been frizzed and last two three layes which are two Fully Connected layer and one softmax layer were tainted with new dataset.Proposed method achived 97(%) Accuracy.In this paper[2] author has used two datasets one for classification to detect covid-19 ,pneumonia and normal classes best accuracy achived was 96.78 (%) using MobileNetV2 with transfer learning.Author have experimented feature extraction based as well as fine tuned based transfer learning.This [6] study experimentally evaluated the application of pretrained deep learning models to classify thyroid histopathology images into two classes namely NT and PTC. Five pre-trained models namely VGG-16, VGG-19, ResNet-50, InceptionV3 and DenseNet-121 were used for this purpose. Two forms of transfer learning (feature extractor and fine tuning) was employed. Experiments were conducted by altering the train-test splits and data augmentation methods. The results showed that DenseNet-121 performed the best on the dataset for both forms of transfer learning..Author[14] have used few shot technique with attention based algorithm to classify the images having very less class representative.In [24]study aims to assess the effectiveness of the most advanced pre-trained model ResNet-50 on the 1000 sample COVID-Chest X-ray dataset.Author [10] have received 96(%) accuracy and 0.98 sensitivity.Author have used DCNN based on ResNet to classify two binary classifications and one multi class classification and achieved the accuracy of 99.9(%),99.8(%) and 97.3(%) respectively from X-Ray images.

The paper [23] suggests that the future of Data Augmentation holds great promise, particularly with the potential use of search algorithms that combine data warping and oversampling techniques. The layered architecture of deep neural networks provides numerous opportunities for implementing Data Augmentation, with the majority of surveyed augmentations operating in the input layer. However, some methods, like DisturbLabel, even manifest in the output layer.The primary focus of the paper [15] is on data augmentation as a solution to enhance the performance of CNNs. Traditional methods, including affine transformations and color modifications, are discussed. The paper then introduces Generative Adversarial Networks (GANs) as a powerful tool for unsupervised image generation, highlighting their applications in various image-related tasks.The paper delves into the concept of texture transfer and style transfer, providing insights into how these methods can be used for data augmentation. The authors propose a fresh perspective on style transfer for data augmentation and demonstrate its application in medical image datasets, specifically in skin lesion analysis, breast histopathology, and breast mammography.This research [21] presents the application of cutting-edge techniques such stacked ensemble models, transfer learning, and artificial neural networks. Combining various convolutional neural network designs to maximize their extraction and classification capabilities is the main concept behind the research. The most dependable classification tool and the best performance are obtained when DenseNet, Xception, and Inception are combined.The summary of the previous work has been listed in table 1

3 Methodology

The research began by computing the base model accuracy for the six selected convolutional neural network (CNN) models: VGG19, VGG16, ResNet50, MobileNetV2, InceptionV3, and DenseNet201. The dataset, encompassing four distinct classes - COVID-19, pneumonia, lung opacity and normal, was divided into training and testing subsets, following standard train-test split ratios. Each CNN model was trained individually on the training data and subsequently assessed using accuracy as the primary metric. This initial phase laid the foundation for evaluating the models' classification performance on the dataset.During the feature extraction phase, all layers of these pre-trained CNN models, except for the output layer, were frozen. Freezing the layers means that the weights and parameters of the earlier layers remain fixed and unaltered. This decision was made to retain the general features learned by these models from their extensive training on large-scale datasets like ImageNet. These general features often encompass basic visual patterns, edges, textures, and shapes that are relevant for a wide range of computer vision tasks, including COVID-19 classification. With the earlier layers frozen, the dataset was passed through the models, and the activations from the final layer before the output layer were extracted. These activations, also referred to as features, represented high-level, abstracted representations of the input images. These features encapsulated relevant information about the dataset, allowing the models to focus solely on learning the fine-grained, task-specific patterns relevant to chest X-ray classification. The extracted features were then used as input to a separate classification layer, typi-

Table 1: Summary of the previous work

Reference	dataset	accuracy	Vrelevance
[7]	3797 X-ray images	accuracy of 97.11(%)	pre-trained VGG-19 architecture
[2]	1427 X-ray images	accuracy of 96.78(%)	CNN with transfer learning
[6]	221thyroid histopathology images	DenseNet-121performed the best for both forms of transfer learning	Two forms of transfer learning(featureextractor and fine tuning)was employed
[10]	5856 CXRimages.	97.3(%) for multi class cases	applied DCNN based on a residual network (Resnet-50)
[24]	datasets having 1000 Chest X-ray	1achieved 96(%) accuracy with 0.98 sensitivity and 0.95 specificity	pre-trained model ResNet-50
[21]	30,000 chest radiographs	1achieved 98(%) accuracy with 0.98 sensitivity and 0.98 specificity	Ensemble model consisting of three different CNN models including DenseNet201, InceptionV3, and Xception.

cally a fully connected neural network, with an appropriate number of neurons and activation functions tailored to the specific classification task. This layer learned to map the extracted features to the distinct classes present in the dataset, namely COVID-19, pneumonia, lung opacity, and normal.Block diagram of feature extraction based transfer learning model as shown in figure 1 . Fine-tuning entails adjusting and optimizing the weights of the pre-trained CNN models to align them more closely with the chest X-ray dataset. Unlike feature extraction, which focuses on using pre-trained features as input, fine-tuning allows for the adaptation of both earlier and later layers of the CNN models. The goal was to refine the models' learned representations, making them more attuned to the distinctive patterns and characteristics of COVID-19, pneumonia, lung opacity, and normal images as shown in figure 2

3.1 Data augmentation

Augment the task-specific dataset with transformations like rotations, flips, and cropping to further enrich the training data.In pursuit of heightened model robustness and the reduction of overfitting, the research introduced a critical component into the training process: data augmentation. This technique was strategically implemented to artificially expand the training dataset by introducing variations and diversity into the existing images. The principles of data augmentation entailed a series of transformations and perturbations applied to the original training images. These transformations included random rotations, horizontal flips, zooming, and shifts, generating augmented versions of each image. The purpose was twofold: firstly, to diminish overfitting risks and ensure that the models' learned representations remained adaptable to diverse data; secondly, to bolster the models' resilience to variations and noise commonly encountered in real-world medical images. Data augmentation, harmoniously integrated with transfer learning in this work, contributed to the development of robust and adaptable classification models, attuned to the complexities of the dataset.

3.1.1 Horizontal flip

One of the employed data augmentation techniques involved horizontal flipping, activated by setting horizontalflip=True in the ImageDataGenerator. This augmentation technique introduced a vital form of variability in the dataset. With a 50(%) chance, it randomly flipped images horizontally, effectively mirroring them. This flip operation altered the orientation of objects within the images. Such variation is crucial for training a robust model capable of recognizing objects from different angles and perspectives.

For a horizontal flip, you reflect the image across the vertical axis. This can be achieved by negating the x-coordinate:

$$x' = -x$$

$$y' = y$$

It enhances the model's generalisation, ensuring that it doesn't overfit to specific orientations presented in the training data. In real-world scenarios, objects can appear in various orientations, and this augmentation method enabled the model to adapt and perform well regardless of the viewpoint.

3.1.2 Rotation range

Another augmentation technique involved rotating the images randomly within a specified range of degrees, controlled by the rotation range parameter. In this research, images were allowed to rotate between -20 and 20 degrees. This augmentation was particularly valuable in scenarios where the training data might lack diversity in object orientations. For rotation, you can rotate the image by an angle

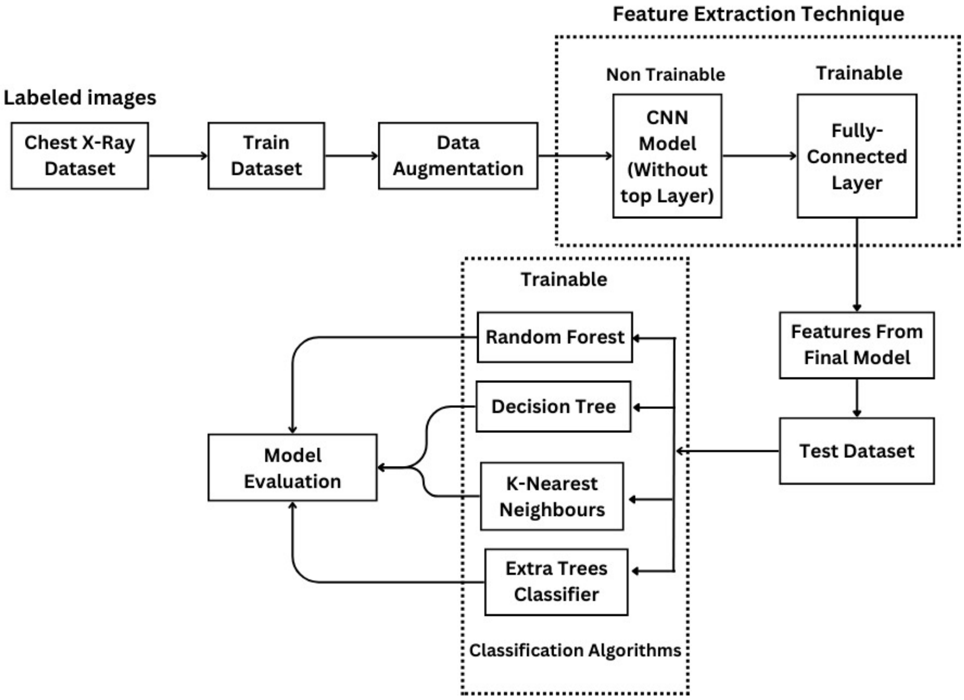

Figure 1: Block diagram of feature extraction based transfer learning model

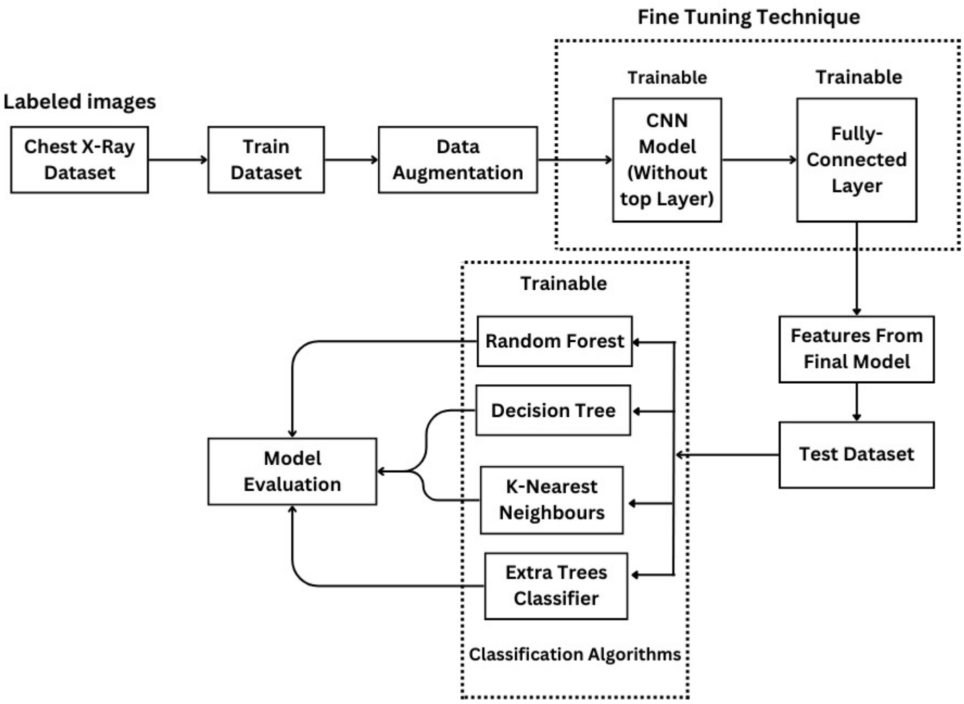

Figure 2: Block diagram of fine tuning based transfer learning model

within a specified range. The rotation angle θ can be determined randomly within the specified range:

$$x' = x \cdot \cos(\theta) - y \cdot \sin(\theta)$$

$$y' = x \cdot \sin(\theta) + y \cdot \cos(\theta)$$

By introducing these rotations, the model learnt to recognise objects at different angles, effectively improving its ability to handle real-world images where objects can appear rotated or tilted. It encouraged the model to develop robust features that were invariant to these variations, ultimately contributing to its classification accuracy.

3.1.3 Zoom range

The third augmentation technique pertained to zooming, implemented via the zoom range parameter. It allowed images to be randomly zoomed in and out by up to 20(%). This augmentation had a notable impact on the model's adaptability. By varying the scale and position of objects within the images, it helped the model learn to detect objects at different sizes. For zooming, you can scale the image by a factor within a specified range. The scaling factor s can be determined randomly within the specified range:

$$x' = x \cdot s$$

$$y' = y \cdot s$$

In real-world scenarios, objects may not always appear at a consistent scale, and this augmentation enhanced the model's capability to handle objects of varying sizes. It encouraged the development of features that were scale-invariant, a crucial aspect of object recognition in practical applications.

3.2 Deep learning

Deep learning is the extention of machine learning which focuses on automatic feature detection and classification .

3.2.1 ResNet-50

ResNet-50 is a widely recognized convolutional neural network architecture that has significantly influenced the field of computer vision [19]. It was specifically designed to address the challenge of training deep neural networks by mitigating the vanishing gradient problem [19][25]. ResNet-50 introduces the concept of residual blocks, which contain skip connections that allow the network to learn residual mappings. These skip connections enable the gradients to flow more directly during backpropagation, making it easier to train very deep networks. By stacking multiple residual blocks, ResNet-50 can learn increasingly complex and discriminative features as the depth of the network increases. This architecture has achieved state-of-the-art performance on various computer vision tasks, including image classification

3.2.2 VGG19

VGG19 is a convolutional neural network architecture that was developed by the Visual Geometry Group (VGG) at the University of Oxford [11], [9], [8]. It is known for its simplicity and effectiveness in capturing fine-grained details and higher-level features . VGG19 follows a straightforward and uniform design philosophy, utilizing multiple convolutional layers with small 3x3 filters and max pooling layers [11].

3.2.3 VGG16

VGG16 is a variant of the VGG network architecture, also developed by the Visual Geometry Group at the University of Oxford [11], [9]. It shares the design philosophy of VGG19, utilizing multiple convolutional layers and max pooling layers for feature learning and extraction. VGG16 has been widely used and influential in the computer vision community. Its simplicity and effectiveness have made it a popular choice for various tasks, including image classification, object detection, and style transfer [11].

3.2.4 InceptionV3

InceptionV3 architecture, also known simply as Inception-ResNet-V3, is a state-of-the-art convolutional neural network (CNN) model designed for image classification and feature extraction tasks [9]. It is an evolution of the original Inception architecture introduced by Google in 2014, which aimed to address the challenges of efficiently processing multi-scale features within a single network

3.2.5 MobileNetV2

MobileNetV2 is an evolution of the MobileNet architecture, specifically designed to address the challenges of deploying deep neural networks on resource-constrained devices, such as mobile phones and edge devices. Developed by Google, MobileNetV2 showcases a remarkable balance between model efficiency and performance, making it a cornerstone in the field of lightweight convolutional neural networks (CNNs).

3.2.6 DenseNet201

DenseNet201, an extension of the DenseNet architecture, embodies a revolutionary approach to convolutional neural network (CNN) design by introducing densely connected layers that foster remarkable feature reuse, model compactness, and accuracy. Traditional CNN architectures connect layers sequentially, leading to the isolation of features learned in earlier layers from those learned later[26]. Dense connections, on the other hand, enable each layer to receive direct input from all previous layers, fostering enhanced feature propagation.

3.3 Transfer learning

Transfer learning is a technique in machine learning where a model trained on one task is used as the starting point for a model on a second task. This can be useful when the second task is similar to the first task or when there is limited data available for the second task. By using the learned features from the first task as a starting point, the model can learn more quickly and effectively on the second task. This can also help to prevent overfitting, as the model will have already learned general features that are likely to be useful in the second task. Many deep neural networks trained on images have a curious phenomenon in common: in the early layers of the network, a deep learning model tries to learn a low level of features, like detecting edges, colors, variations of intensities, etc. Such kinds of features appear not to be specific to a particular dataset or a task because no matter what type of image we are processing . By using a pre-trained model, the model can learn more quickly and effectively on the second task, as it already has a good understanding of the features and patterns in the data.Transfer learning can lead to better performance on the second task, as the model can leverage the knowledge it has gained from the first task. When there is limited data available for the second task, transfer learning can help to prevent overfitting, as the model will have already learned general features that are likely to be useful in the second task. Transfer learning can lead to overfitting if the model is fine-tuned too much on the second task, as it may learn task-specific features that do not generalize well to new data.

3.3.1 Feature extraction

Feature extraction is executed through a series of convolutional and pooling layers in a CNN as shown in Figure 3. Convolutional layers apply filters (also known as kernels) to input images, detecting specific patterns like edges, textures, or gradients. These layers progressively capture increasingly complex features by hierarchically combining lower-level features. Pooling layers downsample the feature maps, reducing their dimensions while retaining essential information. The result is a set of abstracted features that capture different levels of information from the input image.Dimensionality Reduction: By converting raw images into compact feature representations, feature extraction reduces the dimensionality of the data, making it more manageable for subsequent processing steps. Generalization: Extracted features focus on the most distinguishing characteristics of the input data, enhancing the network's ability to generalize across different instances of a class or object. Robustness: Extracted features are often more robust to variations in scale, rotation, lighting conditions, and noise, contributing to the network's robust performance on diverse data [20].

[h!]

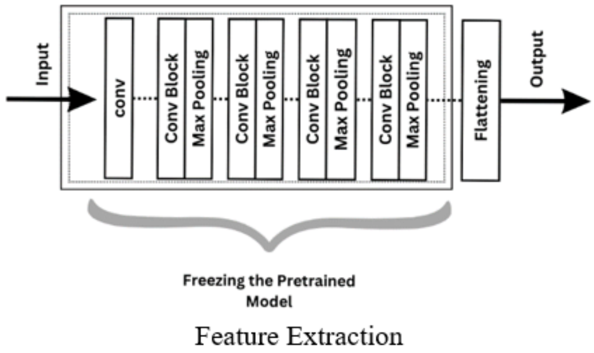

Figure 3: Network diagram of feature extraction based transfer learning

3.3.2 Fine-tuning:

Fine-tuning is a strategic technique in the domain of Convolutional Neural Networks (CNNs) that facilitates the adaptation of pre-trained models to new tasks or domains. By building upon the knowledge learned from a large dataset and transferring it to a more specific task with limited data, fine-tuning strikes a balance between leveraging existing knowledge and tailoring it to new challenges. This process enables deep learning models to achieve higher performance and faster convergence on tasks that share underlying features with the original training data.Fine-tuning involves taking a pre-trained CNN model, often trained on a massive dataset like ImageNet, and adjusting its parameters on a smaller, task-specific dataset. The earlier layers of the pre-trained model capture general features that are applicable across various tasks, while the later layers specialize in extracting task-specific features. Fine-tuning allows the model to retain the general knowledge while refining the higher-level features to better suit the new task.Network diagram of fine-tuning base transfer learning is shown in Figure 4

[h!]

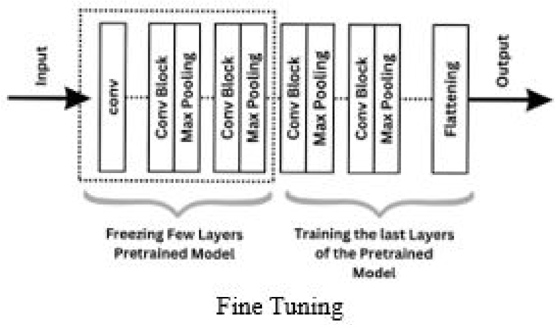

Figure 4: Network diagram of fine tuning based transfer learning

3.4 Classification algorithms

Classification is a fundamental task in machine learning that involves assigning predefined labels or categories to data points based on their features or characteristics. Classification algorithms play a pivotal role in various domains, ranging from image recognition and natural language processing to medical diagnosis and fraud detection. These algorithms are employed to automate decision-making processes, discern patterns, and make predictions in a wide array of applications[25]

3.4.1 Random forest

Random Forest is an ensemble learning algorithm widely used in machine learning for both classification and regression tasks, celebrated for its robustness and high accuracy [22].

3.4.2 Decision tree

It creates a hierarchical tree-like structure that recursively splits the dataset based on features, making decisions at each node to maximize the separation of classes or reduce variance in regression problems. Decision trees are characterized by their simplicity and interpretability [18], allowing users to understand the reasoning behind each decision made by the model.

3.4.3 K-nearest neighbours (KNN)

The K-Nearest Neighbours (KNN) algorithm is a simple yet effective machine learning technique used for both classification and regression tasks [22], [**?**]. It operates on the principle of proximity, where it classifies or predicts the target variable of a data point by considering the majority class or averaging the values of its k-nearest neighbors in the feature space.

3.4.4 Extra trees

The Extra Trees Classifier, short for Extremely Randomized Trees Classifier, is a powerful ensemble learning method within the realm of machine learning and decision trees [**?**]. It builds upon the Random Forest algorithm by introducing an additional level of randomness during the tree construction process.

3.5 Few shot learning

It can be difficult to get a sizable data collection that we can use to train a deep learning model even with transfer learning in many real-world situations.This is made possible by few-shot learning because it allows models to learn from a limited amount of data.In few–shot learning the dateset is divided into two distinct set called support set and query set.K samples of N categories are selected and formed as the support set and rest P samples of N categories forms the query set.

Support Set: $\{x_{i,k} \mid i = 1, 2, ..., N;\ k = 1, 2, ..., K\}$
Query Set: $\{x_{i,p} \mid i = 1, 2, ..., N;\ p = 1, 2, ..., P\}$

A model is given a query sample which belongs to a new unseen class.We will also have one support set consisting of N number of of images of K different class.Calculate the prototype representation for each class by taking the mean of the encoded support set examples for that class as shown in equation 1:

$$\mathbf{c}_i = \frac{1}{K} \sum_{k=1}^{K} \mathrm{Enc}(x_{i,k}) \tag{1}$$

Model then finds which of the support set classes the query sample image belongs to.It actually finds the similarity between the query and support sample in the embedding space [13]. There are two types: one is Prototypical Networks and the another Siamese Networks.Prototypical Networks were introduced by Snell et al. (2017) as a simple yet effective method for few-shot classification. Unlike Siamese Networks, which learn a pairwise similarity function by comparing query and support samples individually, Prototypical Networks compute a single prototype per class and measure distances to these prototypes.

3.6 Dataset

This dataset comprises a total of 3,707 images, categorized into four distinct classes. Specifically, it includes 576 images of COVID-19 cases, 1,052 images of lung opacity, 1066 images of normal chest conditions, and 1,013 images of pneumonia cases. These classes provide essential distinctions for the accurate identification of various respiratory conditions, a critical task in the medical field. Notably, the dataset is a compilation of two renowned datasets which are publicly availabe at kaggle: the ”COVID-19 Radiography Dataset” by Preet Viradiya and the ”Chest X-ray (Covid-19 & Pneumonia)” dataset by Prashant Patel. Every image was shrunk to 224 × 224 pixels, which is the fixed resolution. To align inputs with ImageNet normalization criteria, pixel intensity values were normalized by rescaling them to the [0,1] range. Where necessary, additional preprocessing was undertaken using the preprocess input function. Data augmentation was carried out utilizing zoom transformations, random rotations, and horizontal and vertical flips to improve dataset diversity and decrease overfitting.

3.7 Evaluation metrics

A confusion matrix is a matrix that summarizes the performance of a machine learning model on a set of test data. It is often used to measure the performance of classification models, which aim to predict a categorical label for each input instance. The matrix displays the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) produced by the model on the test data. A confusion matrix has been generated to evaluate the performance of each model. From the confusion matrix, accuracy (ACC) has been calculated to compare the performance.

1. **Accuracy (ACC):** It is the percentage of correctly classified samples.

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

2. **Recall/Sensitivity:** It gauges the percentage of real positives that were accurately named. This also represents the proportion of misdiagnoses.

$$\text{Recall/Sensitivity} = \frac{TP}{TP + FN}$$

Sensitivity measures the frequency with which the model properly identifies a positive COVID-19 example as such.

3. **Precision:** It is the proportion of true positives among detected positives.

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

4. **F1-score:** F1 combines a model's precision and recall ratings. The accuracy statistic counts the number of times a model accurately predicted the whole dataset.

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

4 Experimental analysis

Table 2: Performance of classifier algorithm

Model	Accuracy
RANDOM FOREST	77.43
DECISION TREE	72.43
K NEAREST NEIGHBOURS	79.24
EXTRA TREE CLASSIFIER	71.67

The accuracy of the classifiers is shown in Table2 where the standalone classifiers have been used directly on the raw dataset.Table3 shows the performance of the base model(without any types of transfer learning),the feature extraction based and fine-tuning-based model.From the above mentioned table it is clear that fine-tune-based model provides the best accuracy for all the mentioned deep learning models.Table4 shows time taken for execution of all the mentioned transfer learning models and it shows that the feature extraction based model takes less time as compared to other techniques of transfer learning for all the model.Fine-tune-based transfer learning also has been evaluated by freezing different portions(in percentage) of the network and the accuracy of the same has been displayed in the table 5.

With the foundational groundwork laid through base model accuracy assessment, feature extraction, and fine-tuning, the research proceeded to the crucial phase of classifier integration. The features extracted from the previous stages, serving as highly informative representations of the chest X-ray dataset, were channelled into a range of classifiers. These included well-established algorithms such as Random Forest, Decision Tree, K-Nearest Neighbors (KNN), and Extra Trees Classifier. This integration strategy was designed to capitalize on the comprehensive knowledge and understanding captured in the earlier phases for the precise classification of of the dataset. The feature vectors extracted from the pre-trained CNN models were integral to this phase. These representations, borne out of feature extraction and fine-tuning, encapsulated intricate visual patterns and critical information necessary for accurate classification. These feature vectors became the input data for the machine learning classifiers, thus enabling the classifiers to learn the intricate relationships between the abstracted features and the corresponding labels in the chest X-Ray dataset. The classifiers chosen—Random Forest, Decision Tree, KNN, and Extra Trees Classifier—received these feature vectors as their primary input. They were subjected to rigorous training on this feature data, with their specific parameters optimized for peak classification performance. The assessment of their performance was comprehensive, with a strong focus on accuracy as the primary metric. Additionally, precision, recall and F1-score, score were meticulously calculated for each classifier, providing a holistic view of their classification capabilities. Ensemble learning, a powerful paradigm in machine learning, also played a significant role in this phase. Ensemble techniques like Random Forest and Extra Trees Classifier were considered as potential classifier options. These methods leveraged the aggregated predictions of multiple classifiers to enhance the overall classification performance, capitalizing on the diversity of information contained within the extracted features. The integration of the base model output(without Transfer learning), feature extraction, and fine-tuning with a suite of classifiers represented a pivotal juncture in the research. This integration strategy leveraged the strengths of each phase to enhance the overall accuracy in the classification of the X-ray dataset. It facilitated the transformation of rich, learned feature representations into actionable classification decisions, aligning seamlessly with the research's overarching goal of achieving robust and precise chest X-ray classification.In essence, the integration of these stages brought together deep learning representations with the discriminative power of traditional classifiers, harmoniously contributing to the research's objective of accurate and dependable chest X-ray classification.Table6 shows the compatibility of different classifiers with the techniques of transfer learning.

The three primary impacts (Model, Type, and Classifier) are all statistically significant, as the table7 demon-

Table 3: Accuracy of base model and transfer learning on normal dataset

Model	Base Model Accuracy (%)	Feature Extraction (%)	Fine Tuning (%)
VGG16	88.95	89.35	90.84
VGG19	88.27	95.36	95.55
ResNet50	82.21	95.55	95.55
InceptionV3	85.58	85.98	90.03
MobileNet-V2	87.20	93.13	95.28
DenseNet-201	91.91	94.88	97.44

Table 4: Execution time of base model and transfer learning on normal dataset

Model	Base Model Accuracy	Feature Extraction	Fine Tuning
VGG16	437.94	185.87	339.99
VGG19	565.59	211.29	263.99
ResNet50	458.73	140.95	267.59
InceptionV3	579.34	209.00	271.11
MobileNet-V2	142.81	85.84	152.52
DenseNet-201	695.80	209.55	411.58

strates.Given that different models function differently, the relationship between Model and Type is extremely important.Additionally important is the relationship between the classifier and the model.However, There is no discernible connection between Type and Classifier. The three-way interaction between Model, Type, and Classifier is significant ($p \approx 0.05$), indicating that the specific model architecture determines how the technique and classifier work together. According to the analysis, each of the three criteria affects accuracy, and their total impact is not always additive, particularly for particular models.

In pursuit of heightened model robustness and the reduction of overfitting, the research introduced a critical component into the training process: data augmentation. Figure5 shows the the sample of Xray images after implementing data augmentation.

Table8 Shows the the accuracy comparison of the same deep learning model on the augmented dataset,Table 9 shows the comparison of execution time of all the deep learning models on the augmented dataset.The hardware specification of the machine is RAM : 29GiB GPU T4 x2 : 15GiB.Figure 6 shows the performance of the deep learning model without data augmentation both for base model and Transfer model on different performance matrices.Figure 7 shows the performance of deep learning model after data augmentation, both for the base model and Transfer model on different performance matrices. With the enriched dataset through data augmentation techniques, the research proceeded to further enhance the model's capabilities. The augmented dataset, now imbued with a greater variety of orientations, scales, and perspectives, was employed for training. This augmented dataset became a vital asset in training more robust and adaptable models. The process, which followed the same structure as previously detailed, encompassed base model evaluation, feature extraction, fine-tuning, and classifier integration, with a primary focus on accuracy as the performance metric. The augmentation-infused dataset ensured that the models were well-prepared to handle the complexities and variations present in real-world medical images, making them more dependable for the critical task of chest X-ray classification.

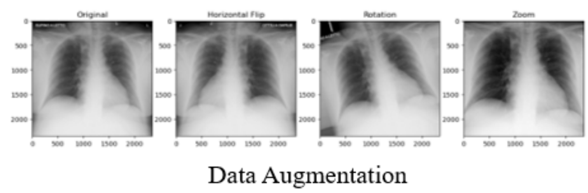

Figure 5: Images after augmentation

These considerations ensured that the resulting chest X-ray classification models not only exhibited high accuracy and robustness but were also practical for real-world implementation, making them valuable tools in medical image analysis and healthcare decision support systems where timely and efficient diagnoses are crucial. Table10 shows the compatibility of different classifiers with the techniques of transfer learning on augmented data model.Table11 depicts the performance of the base model(without transfer learning) and transfer learning on the augmented Dataset-From table 6 and 10 we can say that all the classifiers provide better results on the augmented dataset.A Significant phase of the research involved the application of few-shot learning techniques to a highly constrained dataset. In this scenario, only 15 images from each class of the Chest X-Ray dataset were utilized for training. The integration of different types of few-shot learning into the research marked a crucial step towards addressing the challenge of data scarcity, a common hurdle in medical image analysis. By successfully applying these techniques

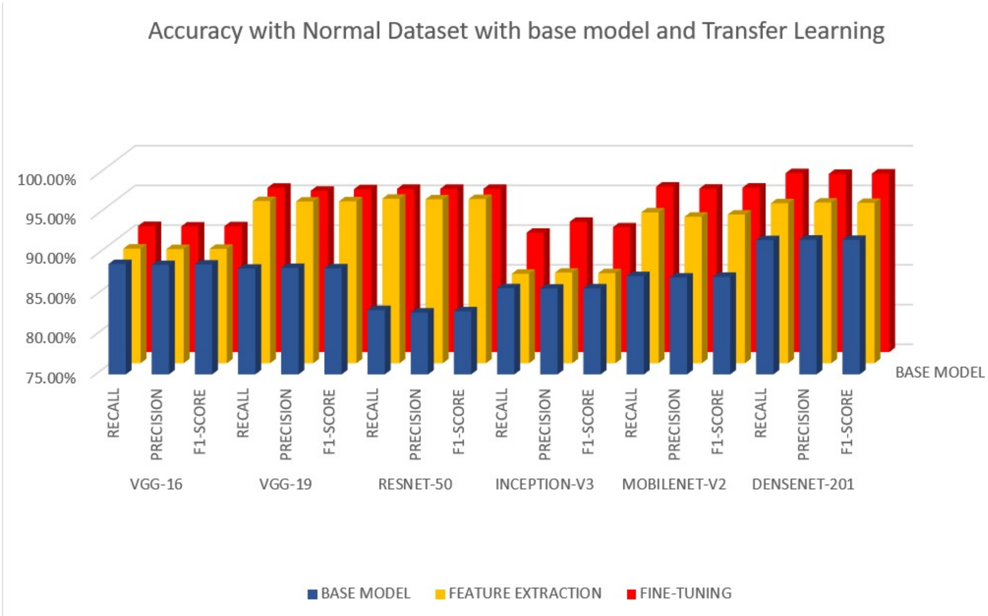

Figure 6: Performance of deep learning model with out data augmentation both for base model and Transfer model

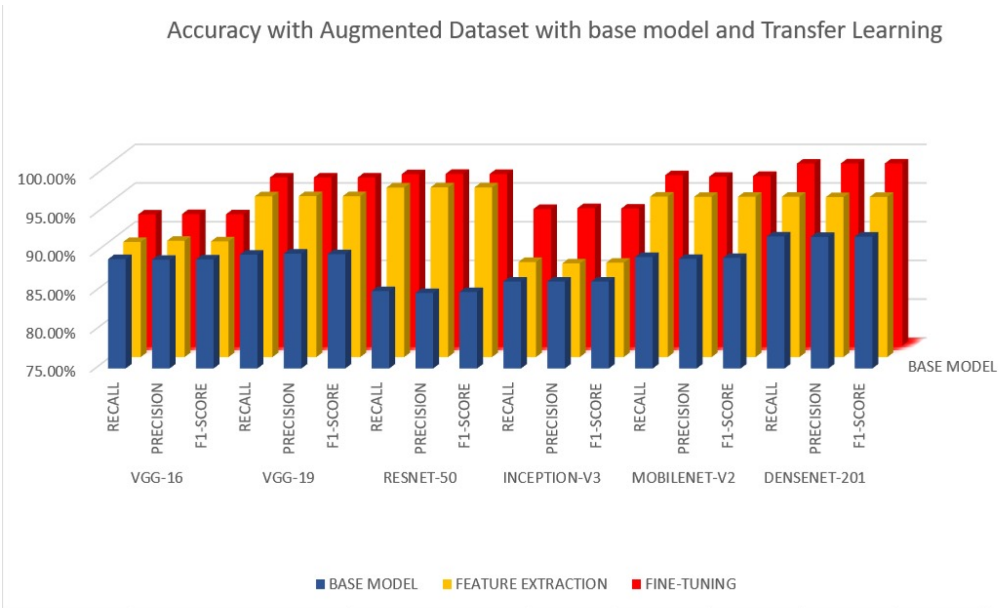

Figure 7: Performance of deep learning model with data augmentation both for base model and Transfer model

Table 5: Accuracy of fine-tune based transfer learning after freezing different portion(percentage) of the model

Model	50 (%)	60 (%)	70 (%)	80 (%)	90 (%)
VGG16	29.65	26.68	28.03	29.25	95.96
VGG19	27.36	27.76	26.55	27.36	95.55
ResNet50	96.36	95.42	95.55	93.13	96.36
InceptionV3	90.70	90.70	86.66	88.41	92.45
MobileNet-V2	54.04	48.79	62.80	26.68	95.28
DenseNet-201	91.57	96.09	95.42	96.23	97.44

Table 6: Compatibility of various classifiers on various models using transfer learning techniques

Model	Type	Random Forest Accuracy(%)	Decision Tree Accuracy(%)	KNN Accuracy(%)	Extra Trees Accuracy(%)
VGG16	Base Model	87.87	86.11	89.08	88.54
VGG16	Feature Extraction	89.74	89.83	89.08	89.49
VGG16	Fine Tuning	90.70	90.02	89.75	90.02
VGG19	Base Model	88.27	87.87	89.08	88.14
VGG19	Feature Extraction	96.22	96.22	95.82	95.68
VGG19	Fine Tuning	95.68	95.68	95.82	96.09
ResNet50	Base Model	90.43	90.16	87.60	90.02
ResNet50	Feature Extraction	90.43	90.16	89.60	90.02
ResNet50	Fine Tuning	95.55	95.55	95.55	95.55
Inception-V3	Base Model	92.58	91.23	91.64	91.64
Inception-V3	Feature Extraction	88.48	87.66	87.94	89.79
Inception-V3	Fine Tuning	92.31	90.43	91.77	92.72
MobileNet-V2	Base Model	89.35	88.27	87.87	88.40
MobileNet-V2	Feature Extraction	95.01	95.01	94.20	94.87
MobileNet-V2	Fine Tuning	95.28	94.87	95.15	95.41
DenseNet-201	Base Model	94.87	93.93	94.74	94.74
DenseNet-201	Feature Extraction	95.14	95.01	95.95	95.82
DenseNet-201	Fine Tuning	97.43	97.57	97.43	97.57

to only 15 images from each class, the research explored innovative ways to make accurate chest X-ray classifications with minimal training samples, potentially opening doors to more efficient and adaptable diagnostic systems in resource-constrained medical environments.Traditional machine learning approaches, including deep learning, often require large datasets to generalize effectively and make accurate predictions.However,in many real-world scenarios, acquiring abundant labelled data for each class is prohibitively expensive,time-consuming, or simply unfeasible.A significant observation arising from the research was the comparative performance of Base model, two distinct learning paradigms: transfer learning() two types) and few-shot learning(two types). These approaches were rigorously evaluated using the limited dataset of 15 images per class from the chest X-ray dataset, with a particular focus on their classification accuracy. This finding underscores the adaptability and robustness of few-shot learning in the context of limited data availability, positioning it as a compelling approach for tasks where comprehensive training datasets are challenging to obtain.The performance of two types of few-short that is Prototypical Networks and Siamese Networks are compared with the transfer learning technique on a reduced dataset(15 images per class) has been shown in table 12.

5 Discussion

The results from the classification algorithms, including Random Forest, Decision Tree, K Nearest Neighbors (KNN), and Extra Trees Classifier, suggest that they have not performed exceptionally well on the dataset. The highest accuracy achieved as shown in Table2 was 79.24(%) with KNN, and the other algorithms had accuracy scores ranging from 71.67(%) to 77.43(%).These results indicate that the dataset may have complex patterns or dependencies that traditional classification algorithms struggle to capture effectively. Therefore, considering the limitations of these algorithms, it appears that exploring the use of Convolutional Neural Network (CNN) models is warranted, particularly if the dataset contains grid-like or image-based data, as CNNs are designed to excel in such scenarios and may yield better classification performance.From Table3 and Table4 we can say that although fine-tune-based model provides the best accuracy among all the models, but it takes more

Table 7: 3-way Anova test on Compatibility of various classifiers on various models using transfer learning techniques

Source	df	F	P Value
Model	5	224.49	< 0.00001
Type (Technique)	2	433.98	<0.00001
Classifier	3	6.43	0.00004
Model and Type	10	86.38	< 0.00001
Model and Classifier	15	2.95	0.00004
Type and Classifier	6	0.6	0.7268
Model,Type and Classifier	30	1.54	0.0496

Table 8: Accuracy of base model and transfer learning on Augmented Dataset

Model	Base Model Accuracy (%)	Feature Extraction (%)	Fine Tuning (%)
VGG16	89.05	89.85	91.84
VGG19	89.84	95.72	96.74
ResNet50	82.75	96.92	97.02
InceptionV3	96.26	87.05	92.63
MobileNet-V2	89.20	96.68	96.86
DenseNet-201	91.98	96.62	98.56

time to execute because some outer layers of the model have been retrained on the actual dataset.The performance of the above mentioned classifiers has been enhanced by taking the features extracted from CNN as input,Table6 shows that Random Forest and Extra Trees Classifier provides better results while they performed ensembling with CNN. Table8 proves that data augmentation have improved the performance of all CNN models for all the specification.From table 6 and and 10 we can say the all the classifiers provide better performance on augmentaed dateset. Table 12 shows that Siamese Networks based Few-Shot provides best accuracy as compaired to two types of transfer learning and prototypical network type few-short learning.From the above study it has been observed that DenseNet 201 is the best CNN model for all the experiments this may be the following reasons.DenseNet201 uses concatenation to link every layer to every layer before it.This implies that gradients and features from previous layers are directly accessible to each layer.It reuses low-level features and cuts down on redundancy, which is particularly useful when training data is limited. Models must be able to learn efficiently from relatively little input in few-shot circumstances.By overcoming the vanishing gradient issue, DenseNet's architecture guarantees improved gradient propagation during backpropagation.Even with a small number of examples, this facilitates the training of deeper models (such as DenseNet201).

6 Limitations and future scope

The primary goal of this work was to handle the data scarcity problem in medical data classification.Experiments can be conducted by taking multimodal data like X-ray images along with CT scan image of the same patient.Study can be further extended by taking medical image along with clinical data of the same patient.

7 Conclusion

In the midst of the ongoing global health crisis, the development of robust and accurate diagnostic tools remains a paramount concern.Our research embarked on a journey exploring adaptability, resilience, and precision in the realm of data scarcity. Commencing with the evaluation of six distinguished convolutional neural network (CNN) models, namely VGG19, VGG16, ResNet50, MobileNetV2, InceptionV3, and DenseNet201. Here base model performance of the above mentioned deep learning models have been compared with the performance of two types of transfer learning ie feature extraction and fine-tuning on chest X-Ray dataset for lung disease classification .The results shows that fine-tuning based model provides the best results as compared to feature extraction based and based model. For fine-tuning, performances have been recoded after freezing different percentage of the model, notably demonstrated peak accuracy of 97.44(%) on DenseNet201 when 90(%)of the model was frozen. Taking a step further, we transferred learned features to classifiers—random forest, decision tree, k nearest neighbors, and extra trees—culminating in an outstanding accuracy of 97.57(%) using DenseNet201 with the fine-tuning technique for decision tree and extra tree classifier. Augmenting the dataset, we observed substantial improvements across experiments. The augmented dataset exhibited higher accuracy than the normal dataset.Fine tuning based DenseNet201 achieved the best accuracy of 98.56(%). Transferring these features to classifiers, particularly using the fine-tuning technique on DenseNet201, yielded an exceptional accuracy of 98.89(%) with extra trees classifiers. In a unique exper-

Table 9: Execution time of base model and transfer learning on augmented dataset

Model	Base Model Accuracy	Feature Extraction	Fine Tuning
VGG16	637.92	412.31	594.49
VGG19	581.25	417.44	462.00
ResNet50	650.93	411.47	436.04
InceptionV3	838.58	670.60	805.55
MobileNet-V2	421.20	360.69	389.85
DenseNet-201	726.73	626.92	472.43

Table 10: Compatibility of various classifiers on various models using transfer learning techniques after using data augmentation

Model	Type	Random Forest Accuracy(%)	Decision Tree Accuracy(%)	KNN Accuracy(%)	Extra Trees Accuracy(%)
VGG16	Base Model(without Transfer learning)	89.62	88.92	89.15	88.97
VGG16	Feature Extraction	90.54	90.05	89.92	90.42
VGG16	Fine Tuning	91.89	91.57	91.92	92.14
VGG19	Base Model	89.80	89.95	89.76	89.21
VGG19	Feature Extraction	96.37	96.42	95.98	96.28
VGG19	Fine Tuning	96.75	97.24	96.41	96.38
ResNet50	Base Model	93.41	92.15	93.22	92.84
ResNet50	Feature Extraction	96.96	96.43	95.64	96.99
ResNet50	Fine Tuning	97.15	97.22	96.95	97.19
Inception-V3	Base Model	92.58	91.23	91.64	91.64
Inception-V3	Feature Extraction	88.31	88.08	88.57	88.63
Inception-V3	Fine Tuning	92.74	92.23	92.48	92.81
MobileNet-V2	Base Model	89.42	88.73	89.46	89.52
MobileNet-V2	Feature Extraction	95.61	96.24	95.63	96.59
MobileNet-V2	Fine Tuning	96.75	96.94	96.71	97.28
DenseNet-201	Base Model	95.41	94.34	94.96	95.05
DenseNet-201	Feature Extraction	95.74	96.75	96.45	96.27
DenseNet-201	Fine Tuning	98.79	98.64	98.31	98.89

iment with a dataset featuring only 15 images per class, Siamese Networks few-shot learning emerged as a valuable approach, outshining both base models and traditional transfer learning techniques with an accuracy of 96.84(%). Our research showcases the robustness of transfer learning, the efficacy of data augmentation, and the promise of few-shot learning in overcoming data scarcity challenges. These findings not only advance our understanding of model adaptability and precision but also offer practical avenues for optimizing performance in computer vision tasks, laying a foundation for future explorations in this dynamic field.

References

[1] Alvar Agusti, Claus F Vogelmeier, and David MG Halpin. Tackling the global burden of lung disease through prevention and early diagnosis. *The Lancet Respiratory Medicine*, 10(11):1013–1015, 2022. https://doi.org/10.1016/s2213-2600(22)00302-2.

[2] Ioannis D Apostolopoulos and Tzani A Mpesiana. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Physical and engineering sciences in medicine*, 43:635–640, 2020. https://doi.org/10.1007/s13246-020-00865-4.

[3] VS Aruna and J Vijayashree. A critical analysis of brain tumor mri segmentation and classification utilizing machine learning and deep learning methods. *Informatica*, 49(24), 2025. https://doi.org/10.31449/inf.v49i24.8202.

[4] Andrea Asperti and Claudio Mastronardo. The effectiveness of data augmentation for detection of gastrointestinal diseases from endoscopical images. *arXiv preprint arXiv:1712.03689*, 2017. https://doi.org/10.5220/0006730901990205.

Table 11: Performance of base model and transfer learning on augmented dataset

Model	Metrics	Base Model	Feature Extraction	Fine tuning
VGG16	RECALL	89.14%	89.88%	91.96%
VGG16	PRECISION	89.06%	90.04%	92.01%
VGG16	F1-SCORE	89.10%	89.96%	91.98%
VGG19	RECALL	89.72%	95.78%	96.74%
VGG19	PRECISION	89.85%	95.81%	96.74%
VGG19	F1-SCORE	89.78%	95.79%	96.74%
ResNet50	RECALL	85.01%	96.92%	97.17%
ResNet50	PRECISION	84.76%	96.94%	97.21%
ResNet50	F1-SCORE	84.88%	96.93%	97.19%
INCEPTION-V3	RECALL	86.24%	87.29%	92.67%
INCEPTION-V3	PRECISION	86.24%	87.11%	92.78%
INCEPTION-V3	F1-SCORE	86.24%	87.20%	92.72%
MOBILENET-V2	RECALL	89.41%	95.74%	97.04%
MOBILENET-V2	PRECISION	89.17%	95.71%	96.84%
MOBILENET-V2	F1-SCORE	89.29%	95.72%	96.94%
DENSENET-201	RECALL	92.07%	95.73%	98.54%
DENSENET-201	PRECISION	92.01%	95.68%	98.56%
DENSENET-201	F1-SCORE	92.04%	95.70%	98.55%

Table 12: After using few-shot learning

Model	Base Model Accuracy	Feature Extraction	Fine tuning	Few Shot Accuracy (prototypical network)	Few Shot(Siamese Networks)
VGG16	78.62	81.54	82.52	90.62	92.87
VGG19	78.27	79.36	82.51	91.64	93.42
ResNet50	69.21	72.61	81.56	89.47	91.83
InceptionV3	71.24	71.48	73.43	92.25	94.02
MobileNet-V2	77.46	68.85	81.29	91.64	92.96
DenseNet-201	81.15	82.72	85.46	95.22	96.84

[5] Pranjal Bhardwaj, Thejineaswar Guhan, and BK Tripathy. Computational biology in the lens of cnn. In *Handbook of Machine Learning Applications for Genomics*, pages 65–85. Springer, 2022. https://doi.org/10.1007/978-981-16-9158-4_5.

[6] Vijaya Gajanan Buddhavarapu et al. An experimental study on classification of thyroid histopathology images using transfer learning. *Pattern Recognition Letters*, 140:1–9, 2020. https://doi.org/10.1016/j.patrec.2020.09.020.

[7] Soarov Chakraborty, Shourav Paul, and KM Azharul Hasan. A transfer learning-based approach with deep cnn for covid-19-and pneumonia-affected chest x-ray image classification. *SN Computer Science*, 3:1–10, 2022. https://doi.org/10.1007/s42979-021-00881-5.

[8] Mohamed Elgendy. *Deep learning for vision systems*. Simon and Schuster, 2020.

[9] R Gayathri, T Gobinath, A Muthumari, and RSV Swathi. Enhanced ai based feature extraction technique in multimedia image retrieval. *ICTACT Journal on Image & Video Processing*, 13(4), 2023. https://doi.org/10.21917/ijivp.2023.0429.

[10] Fatima-Zohra Hamlili, Mohammed Beladgham, Mustapha Khelifi, and Ahmed Bouida. Transfer learning with resnet-50 for detecting covid-19 in chest x-ray images. *Indonesian Journal of Electrical Engineering and Computer Science*, 25(3):1458–1468, 2022. https://doi.org/10.11591/ijeecs.v25.i3.pp1458-1468.

[11] Shardul Jeurkar, Tanmay Devare, and Karan Borkar. Apple tree disease detection using vgg16 & inceptionv3. 2023. https://doi.org/10.21203/rs.3.rs-3111729/v1.

[12] Bipal Khanal, Astha Singh, Sourav Paul, and Ranjita Das. Multi-class classification on chest x-ray images using convolution neural network. In *2021 IEEE 18th India Council International Conference (INDICON)*, pages 1–5. IEEE, 2021. https://doi.org/10.1109/indicon52576.2021.9691747.

[13] Xiaoxu Li, Xiaochen Yang, Zhanyu Ma, and Jing-Hao Xue. Deep metric learning for few-shot image classification: A selective review. *arXiv e-prints*, pages arXiv–2105, 2021.

[14] Xiaoxia Meng, Xiaowei Wang, Shoulin Yin, and Hang Li. Few-shot image classification algorithm based on attention mechanism and weight fusion. *Journal of Engineering and Applied Science*, 70(1):1–14, 2023. https://doi.org/10.1186/s44147-023-00186-9.

[15] Agnieszka Mikołajczyk and Michał Grochowski. Data augmentation for improving deep learning in image classification problem. In *2018 international interdisciplinary PhD workshop (IIPhDW)*, pages 117–122. IEEE, 2018. https://doi.org/10.1109/iiphdw.2018.8388338.

[16] S. Paul, S. Agarwal, and R Das. Detection of covid-19 using resnet on ct scan image. *Proceedings of International Conference on Computational Intelligence, Data Science and Cloud Computing. Lecture Notes on Data Engineering and Communications Technologies, vol 62. Springer, Singapore.*, 2021.

[17] Ramaprasad Poojary, Roma Raina, and Sachidananda Krishanmurthy. Application of cnns in home security. In *2022 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, pages 322–327. IEEE, 2022. https://doi.org/10.1109/icecta57148.2022.9990490.

[18] D Rajeswari R Srinivasan. A framework for classifying imbalanced tweets using machine learning techniques. *Perspectives on Social Welfare Applications' Optimization and Enhanced Computer Applications*, 2023. https://doi.org/10.4018/978-1-6684-8306-0.ch001.

[19] S Raghavendra, Divya Rao, SK Abhilash, Venu Madhav Nookala, and Praveen Gurunath Bharathi. Elevating amodal segmentation using ash-net architecture for accurate object boundary estimation. *IEEE Access*, 2023. https://doi.org/10.1109/access.2023.3301724.

[20] Aswiga RV and Shanthi AP. Augmenting transfer learning with feature extraction techniques for limited breast imaging datasets. *Journal of Digital Imaging*, 34(3):618–629, 2021. https://doi.org/10.1007/s10278-021-00456-z.

[21] Ruaa Sadoon and Adala Chaid. Classification of pulmonary diseases using a deep learning stacking ensemble model. *Informatica*, 48(14), 2024. DOI: 10.31449/inf.v46i1.4659.

[22] S Selvakani, K Vasumathi, and V Aadhiseshan. Application of machine learning in predicting heart disease. *Asian Basic and Applied Research Journal*, pages 61–68, 2023.

[23] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48, 2019. https://doi.org/10.1186/s40537-019-0197-0.

[24] V Sreejith and Thomas George. Detection of covid-19 from chest x-rays using resnet-50. In *Journal of Physics: Conference Series*, volume 1937, page 012002. IOP Publishing, 2021. https://doi.org/10.1088/1742-6596/1937/1/012002.

[25] Quang Ngoc The Ho, Thanh Trung Do, Pham Son Minh, Van-Thuc Nguyen, and Van Thanh Tien Nguyen. Turning chatter detection using a multi-input convolutional neural network via image and sound signal. *Machines*, 11(6):644, 2023. https://doi.org/10.3390/machines11060644.

[26] Shui-Hua Wang and Yu-Dong Zhang. Densenet-201-based deep neural network with composite learning factor and precomputation for multiple sclerosis classification. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 16(2s):1–19, 2020. https://doi.org/10.1145/3341095.