ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 5, May 2024

Early Detection of Interstitial Lung Disease (ILD)

[1] Nishanth Artham, [2] Rithesh Kunta, [3] Ramakrishna Kolikipogu, [4] K.H. Vijaya Kumari*

[1] [2] [3] [4] Department of Information Technology, Chaitanya Bharathi Institute of Technology (CBIT), Hyderabad, India

Corresponding Author Email: [1] arthamnishanth123@gmail.com, [2] ritheshbinnu09@gmail.com, [3] ramakrishna it@cbit.ac.in, [4] vijayakumari it@cbit.ac.in

Abstract— ***Interstitial Lung Disease (ILD) is a prevalent and progressive respiratory condition that imposes a significant global health burden. Accurate diagnosis of ILD is vital for effective management and intervention strategies. In this paper, we present a novel classification approach for ILD using a Vision Transformer (ViT) model. Vision Transformer (ViT) presents an innovative deep learning framework that has consistently showcased exceptional performance across a wide range of computer vision applications. We aim to explore its applicability in the domain of medical imaging for ILD diagnosis. The proposed method leverages a dataset comprising chest X-rays and CT scans from patients with ILD, as well as healthy controls. Our methodology incorporates self-attention mechanisms to capture long-range dependencies within the images. This enables the model to effectively discern relevant patterns and features. By fine-tuning the pre-trained ViT model on this dataset, we employ transfer learning to adapt the model for the specific task of ILD classification. Our objective is to attain a notable accuracy without a substantial increase in parameter usage during the fine-tuning process. Accurate early-stage diagnosis of ILD through non-invasive imaging techniques holds the potential for timely interventions and improved patient outcomes. Our findings underscore the capability of Vision Transformers as a potent tool in medical image analysis. This research paves the way for enhanced diagnostic capabilities in the field of respiratory medicine.***

Index Terms— ***Vision Transformer, Deep learning, High Resolution Computed Tomography (HRCT) scans, Medical Imaging, Interstitial Lung Disease (ILD).***

I. INTRODUCTION

Lung diseases [1] pose a formidable challenge to global public health, with their prevalence steadily increasing due to various factors, including smoking, environmental pollution, and an aging population. The insidious nature of many lung diseases, often presenting with minimal or no symptoms until they reach advanced stages, underscores the critical importance of early detection methods. Effective early detection strategies, such as screening programs and heightened awareness campaigns, are essential to address this pressing public health concern and enhance patient outcomes.

Medical imaging research mainly focuses on determine the presence of interstitial lung disease (ILD). [2] ILD encompasses a diverse array of disorders characterized by inflammation and fibrosis of the lung’s interstitial tissue, leading to compromised lung function and a diminished quality of life. The primary focus of this research is to employ state-of-the-art classifiers to detect ILD, by utilizing Computed Tomography (CT) scans of the lungs.

The three main objectives of this paper are:

1) To provide a detailed examination of the current classification models on ILD.
2) To elucidate the potential of CT scans in the detection of ILD, highlighting their role in identifying subtle abnormalities and facilitating timely intervention.
3) To discuss the emerging role of vision transformers, originally designed for image classification tasks, in the realm of medical imaging, particularly their application in the detection of ILD.

(a) Vision Transformers (ViTs):

Computer vision has witnessed a groundbreaking paradigm shift with the emergence of Vision Transformers (ViTs) [3]. While convolutional neural networks (CNNs) have long been the cornerstone of image analysis tasks, Vision Transformers represent a transformative approach to image understanding and classification. At its core, a Vision Transformer leverages the selfattention mechanism, allowing it to weigh the importance of different parts of an image when making predictions. This mechanism enables Vision Transformers to process images holistically, considering the relationships between all pixels or patches rather than relying solely on local features as in traditional CNNs [4]. As a result, Vision Transformers excel at capturing complex and context-rich information within images, making them well-suited for a wide range of computer vision tasks.

(b) CT Scans:

Medical imaging has undergone a remarkable evolution, revolutionizing the way healthcare professionals diagnose and manage a spectrum of diseases. Among the myriad of imaging modalities, Computed Tomography (CT) scans have emerged as vital in the realm of radiology and clinical diagnostics. [5] CT scans provide an unparalleled window into the human body, offering detailed, cross-sectional images of internal structures with exceptional spatial resolution. In this paper, we underscore the paramount importance of CT scans in the ILD detection, characterization, and management of a diverse array of medical conditions.

Lung diseases, including ILD, are a matter of significant concern to healthcare professionals, researchers, and policymakers. By synthesizing existing research findings, clinical insights, and cutting-edge technological advancements, this research aims to serve as a valuable resource for those involved in the management and comprehension of ILD. Ultimately, the knowledge presented herein can contribute to the enhancement of patient care and the amelioration of outcomes for individuals grappling with the complexities of ILD.

The subsequent sections of this paper will delve into the multifaceted landscape of ILD, exploring its diagnostic approaches, CT scan image exploration, preprocessing of CT scan images and usage of state-of-the-art classifier vision transformer on ILD database. Furthermore, we will elucidate the promising role of vision transformers and their potential to augment diagnostic accuracy and efficiency. In doing so, we aspire to highlight the significance of innovative technologies in the field of medical imaging, with a specific focus on the realm of lung disease detection..

II. RELATEDWORK

Lung diseases are prominent health issues that affect millions of people worldwide. These diseases can range from acute respiratory infections, such as the common cold and influenza, to chronic conditions like chronic obstructive pulmonary disease (COPD), as well as more severe and lifethreatening illnesses such as lung cancer and interstitial lung diseases (ILDs). The two factors leading lung diseases to become a prominent disease are high prevalence and leading causes of death. Lung diseases are prevalent globally, affecting people of all ages. Respiratory infections like pneumonia and bronchitis are common, especially in children and the elderly contributing to high prevalence. There are three primary categories of lung diseases [6], namely Airway diseases, Lung tissue disorders, and Lung circulation conditions. This article focuses on Interstitial Lung Diseases (ILD), a specific subset of lung ailments that primarily impact the interstitial tissue of the lungs, which surrounds and supports the air sacs (alveoli). ILD encompasses over 200 distinct lung disorders, all of which share common characteristics like lung tissue inflammation and scarring, known as fibrosis. Some well-known ILDs include idiopathic pulmonary fibrosis (IPF), sarcoidosis, and ILD associated with connective tissue diseases.

Medical imaging is of utmost importance in the diagnosis and treatment of ILD, as it involves capturing images of the body’s internal structures for medical purposes. In their study, H. Mary Shyni et al. [7] emphasized the significance of CT and X-ray scans in examining the lung interstitium. Medical imaging encompasses a range of techniques, such as X-ray imaging, computed tomography (CT), nuclear medicine, and ultrasound, to aid in patient care. Image processing techniques are applied in medicine for various purposes, including segmentation and texture analysis for disease identification, image registration, and fusion, telemedicine, and compression for remote image communication. [8] [9] uses MRI images to detect diseases. [10] uses MRI images to detect brain tumour It discusses the use of deep learning models, including IVX16, for multiclass classification of brain tumors in MRI images, aiming to improve accuracy and reliability compared to traditional methods, with a focus on explaining model performance and exploring Vision Transformer (ViT) models. Another medical images are xrays. [11] uses X - Ray images to classify different Lung Disease This paper presents different preprocessing methods like Intensity Normalisation, Gaussian Filter and few data Augmentation methods. The last kind of imaging technique is CT or HRCT scans. [4] analyzes interstitial lung diseases from CT images. The proposed lightweight U-Net architecture maintained segmentation performance compared to the original U-Net while being more computationally efficient.

ContrastLimited Adaptive Histogram Equalization (CLAHE) is an image processing technique that is used to enhance the contrast and improve the visual quality of digital images. It is particularly useful when dealing with images that have variations in illumination and contrast across different regions. CLAHE works by dividing the image into smaller blocks or tiles and then applying a histogram equalization technique to each block individually. This adaptive approach ensures that the local contrast in each region of the image is improved while preventing the amplification of noise or artifacts in the process. By limiting the contrast enhancement, CLAHE prevents the over-amplification of noise in regions with low contrast. Janan Arslan et al in [12] applied CLAHE preprocessing technique for lesion segmentation.

Training the model with a diverse dataset improved its ability to cope with severe pathological regions. Francisco Silva et al in [13] study highlights the need for diverse and representative datasets to build robust segmentation tools for lung analysis. This particular paper discusses in depth about the ILD but few points should be improved like Tumor regions were sometimes included in lung mask predictions due to differences in contouring guidelines among datasets. This needs improvement for accurate tumor inclusion. The lightweight architecture’s performance was evaluated mainly on publicly available datasets. Further validation on a wider range of datasets and clinical scenarios is needed. Future work could focus on refining the model’s performance for accurately segmenting tumor regions and exploring more advanced architectures to address specific challenges in lung CT segmentation. [14] proposes lung ultrasound surface wave elastography (LUSWE) a technique of lung ultrasound surface wave elastography for measuring lung tissue stiffness. ILD occurs when the lung’s interstitial tissue is affected. ILD is a collection of more than 200 diseases and one specific disease is discussed in [2] a comprehensive guide on connective tissue disease-associated interstitial lung disease (CTDILD), developed through collaboration between respiratory and rheumatology experts, aiming to provide evidence-based recommendations and address the heterogeneous nature of CTDILD for improved patient care and future research. Hossain et al in [15] employed vision transformer and other deep learning techniques on MRI images to detect brain tumour. To detect ILD we are proposing [16] Vision transformer(ViT) It is built upon the transformer architecture, which was initially developed for sequence-to-sequence tasks in NLP. The transformer architecture introduced the concept of selfattention mechanisms, which allows the model to weigh the importance of different parts of the input sequence when making predictions. This self-attention mechanism is a crucial component of Vision Transformers. Its components include Image Patching, Positional Encoding, Tokenization, MultiHead Self-Attention, Feed-Forward Layers, Layer Normalization, Output Layer . [17] Vision Transformers are often pre-trained on large datasets, such as ImageNet, using selfsupervised learning techniques. After pre-training, they can be fine-tuned on smaller, task-specific datasets for specific computer vision tasks, such as object detection or image segmentation.

III. MATERIAL AND METHODOLOGY

A. Dataset Description

The research dataset employed in this study comprises a collection of CT scans of lungs in DICOM format [18] obtained from individuals diagnosed with Interstitial Lung Disease (ILD). These images are converted from DICOM to png format as gray-scale images. ILD encompasses a diverse group of lung disorders characterized by inflammation and fibrotic scarring within the lung interstitium. The dataset primarily focuses on three specific ILD subtypes, namely Interstitial Pneumonia, Adenocarcinoma, and Scleroderma. In addition to these ILD-afflicted cases, the dataset also includes CT scans of lung images from individuals who do not exhibit any signs of ILD, representing the control or normal group in this research. The dataset is divided with 70% data for training, 10% for validation, and 20% for testing purposes.

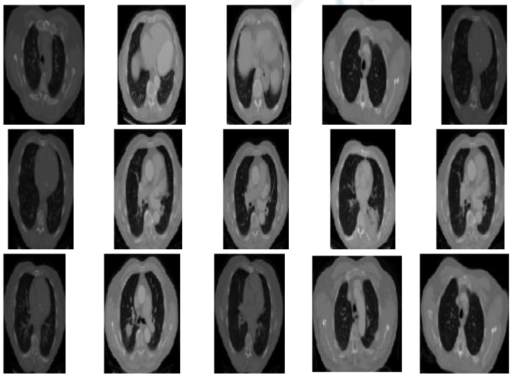

Fig. 1. Workflow of proposed methodology

B. Image Preprocessing

1) Image Enhancement:

- Intensity Normalisation: We employed the Min-Max Scaling Intensity Normalization preprocessing technique

[19] to transform pixel intensities within a specified range. In the context of lung gray-scale images, this technique ensures that all images within a dataset share a consistent intensity scale. Through this technique, pixel intensities are confined within the minimum (0) and maximum (1) values, effectively standardizing the intensity scale.

The Min-Max Scaling formula is expressed as:

$$N_p = \frac{O_p - M_i}{M_i - m_i}$$

where

N_p represents the Normalized pixel value.

O_p denotes the Original pixel value. M_i is the maximum intensity, which is set to 1. m_i is the minimum intensity, which is set to 0.

- Contrast Enhancement

The inherent challenge in medical imaging lies in the variations in image contrast and quality. Inadequate contrast and visibility of anatomical structures can hinder the diagnostic process. Contrast enhancement improves the visual quality and interpretability of images. It aims to increase the distinction between different regions or structures within an image, making them more discernible. ILD-related abnormalities, such as fibrotic changes and ground-glass opacities, can be subtle and challenging to identify in radiological images. Histogram Equalization can reveal these abnormalities with greater clarity, improving diagnostic accuracy. Figure 4 visually illustrates the profound impact of Histogram Equalization as an image preprocessing technique for the detection of Interstitial Lung Disease (ILD). ILD detection demands a meticulous analysis of lung images, as even subtle abnormalities in the interstitial tissue can hold critical diagnostic information. In our study, we employed Histogram Equalization to enhance image contrast, thereby revealing intricate details that might otherwise remain hidden. The histogram comparison graph showcases the transformation of pixel intensity values before and after Histogram Equalization. The elevation in the graph represents the magnitude of the increase in pixel intensity, a crucial factor in identifying scarring and anomalies within the interstitial tissue.

- Resizing and Cropping In addition to Histogram Equalization, we employed resizing and cropping as essential preprocessing techniques to further optimize the CT images for ILD detection.

Resize: One of the initial steps in our preprocessing pipeline involves resizing the medical images. Through this images are uniformly adjusted to a specified dimension, ensuring consistency in size. In our study, we resized the images to a dimension of 256x256 pixels. This not only simplifies the computational load but also enables compatibility with neural networks that require a fixed input size.

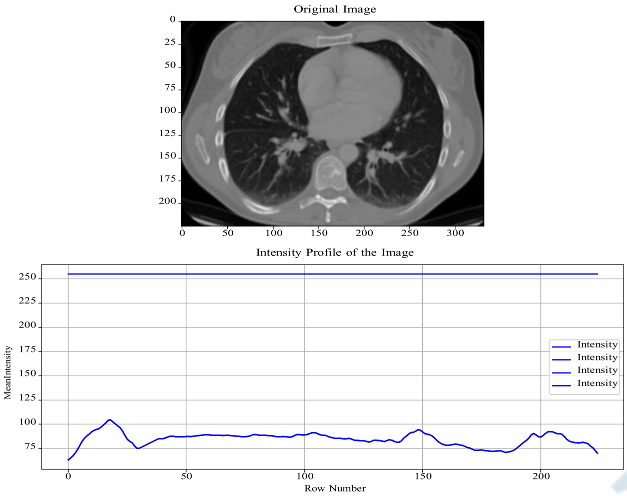

Fig. 2. Before applying Histogram Equalization

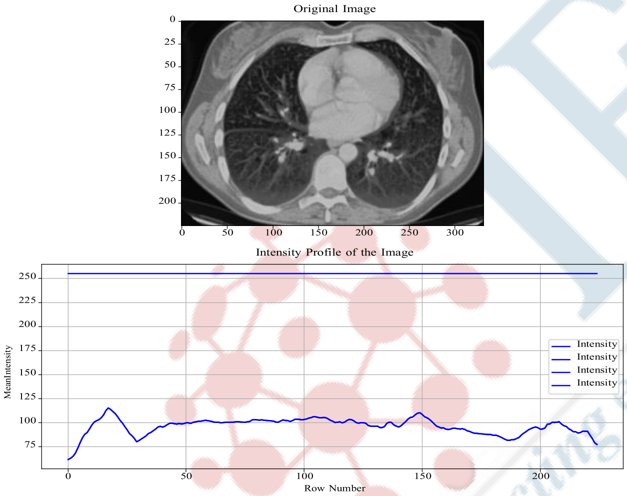

Fig. 3. After applying Histogram Equalization
Fig. 4. Image Intensity graph

Cropping: Cropping is crucial for isolating the regions of interest within the images and eliminating unnecessary information from the periphery. By center-cropping the resized images to a dimension of 224x224 pixels, we concentrate on the core features and structures that are most relevant to ILD detection.

Resizing and cropping ensure that images adhere to a standardized format, facilitating the subsequent analysis and interpretation of ILD-related patterns. By combining these techniques with Histogram Equalization, we offer a comprehensive approach to image preparation that is geared towards optimizing the performance of ILD detection models.

2) Data Augmentation:

Given the limited availability of ILD images in our dataset, data augmentation emerged as a crucial strategy to amplify the diversity of training samples and enhance the robustness of our ILD detection model.

- Horizontal and Vertical Flipping: Flipping operations are fundamental invariance augmentations that simulate changes in the orientation of the patient. By applying horizontal and vertical flips, the model learns to recognize features and patterns regardless of their spatial orientation. These augmentations contribute to improved generalization, making the model less sensitive to image orientation and better equipped to handle images captured in different settings. Therefore, the model becomes spatial invariant.
- Color Jittering: ColorJitter operation introduces random variations in image color and brightness, replicating variations in illumination and color balance frequently encountered in real-world medical images. This augmentation technique enhances the model’s adaptability to varying imaging conditions, ultimately making the model invariant to variations in image color and brightness.
- Random Erasing: It is an effective technique to simulate occlusions and artifacts within the images. By randomly removing portions of the image, the model becomes more robust to the presence of unexpected noise or clutter in the medical images. This augmentation can help the model discern relevant information from potentially distracting elements, thus enhancing its diagnostic accuracy.

The total increase in the effective amount of training data is primarily due to horizontal and vertical flips. Overall, the percent increase in the effective amount of training data is approximately 200% compared to the original data.

C. Classification

Before performing classification techniques the preprocessed data has to be divided. We divided our dataset into three distinct parts: training, validation, and testing. This division was based on a 70-20-10 rule.

Training Phase: The training phase serves as the foundational stage of our model’s learning journey. During this phase, the model is exposed to a substantial portion of the data. The process begins with the model making predictions based on its current understanding of the data. These predictions are then compared to the ground truth labels in the training data, resulting in a measure of how well or poorly the model is performing. This measure, often referred to as ”loss,” is used to update the model’s internal parameters in a way that minimizes the error between predictions and actual labels. This is repeated for many rounds (known as epochs) until the model converges to a state where it accurately recognizes ILD in the images.

$$L_{hardDistill,global} = \frac{1}{2}L_{CE}(\psi(Z_s), y) + \frac{1}{2}L_{CE}(\psi(Z_s), y_{hard}) \quad (1)$$

Validation Phase: The validation phase plays a pivotal role in monitoring the model's performance and preventing overfitting. Here, a separate dataset (the validation set) is used to evaluate the model's accuracy. The model's performance on this validation set helps us fine-tune its parameters and hyperparameters, ensuring that it generalizes well to new, unseen data.

Testing Phase: Finally, the testing phase assesses the model's readiness for real-world application. It is completely independent of the training and validation data. The model's performance is rigorously evaluated on the testing set, providing an unbiased measure of how well it can identify ILD in completely new images.

Classification Models: To detect Interstitial Lung Disease (ILD), we used deep learning and transfer learning techniques.

1) Deep Learning with ResNet50 and VGG16: We employed ResNet50 and VGG16, two well-known deep learning models, to help us classify lung images. While they performed decently, they didn't quite reach the level of accuracy we were aiming for.
2) Transfer Learning with DeiT-B: For even better results, we delved into transfer learning. We took a pre-trained model called DeiT-B (Data-Efficient Image Transformers) [20] and fine-tuned it for ILD detection. This involved keeping the model's foundational layers frozen and then adding our own layers, like dropout and a linear layer, to handle the specific ILD classification task. The DeiT-B model, with this fine-tuning, showed great promise and delivered impressive results. The DeiT model is a convolution-free neural network architecture that uses self-attention mechanisms to process image data. Unlike traditional convolutional neural networks (convnets), which use convolutional layers to extract features from the input image, the DeiT model uses self-attention mechanisms to attend to different parts of the image and learn spatial relationships between them.

In the DeiT model, the input image is first divided into a set of non-overlapping patches, which are then flattened and fed into the model as a sequence of tokens. Each token is associated with a learnable embedding vector, which is used to represent the corresponding image patch. The model then applies a stack of transformer blocks to the sequence of tokens, where each block contains a multi-head self-attention mechanism and a feedforward neural network. The self-attention mechanism allows the model to attend to different parts of the input image, while the feedforward network applies non-linear transformations to the attended features. The output of each transformer block is passed through a layer normalization and residual connection before being fed into the next block.

By using self-attention mechanisms instead of convolutional layers, the DeiT model is able to capture longrange dependencies between different parts of the image, without being limited by the size of the convolutional kernel. This allows the model to learn more complex and abstract representations of the input image, leading to improved performance on image understanding tasks. Additionally, the DeiT model is more computationally efficient than convnets, since it does not require expensive convolutional operations.

In this classification we used hard label distillation. It is a variant of knowledge distillation where the model is trained to match the hard decisions, rather than the full probability distribution. This is done by using the hard decision of the ground truth as a true label. The formula for the distillation objective function for hard label distillation is given by:

where L_{CE} is the cross-entropy loss, ψ is the softmax function, Z_s is the logits of the student model, y is the ground truth label, and y_{hard} is the hard decision of the teacher model.

The multi-head self-attention mechanism allows the transformer to attend to different parts of the input image and capture long-range dependencies between patches. This is achieved by computing multiple attention heads in parallel, each with its own set of learned parameters. The outputs of these attention heads are concatenated and projected back to the original embedding dimension, resulting in a more expressive representation of the input image.

By using the multi-head self-attention mechanism instead of CNNs, the vision transformer is able to achieve state-of-the-art performance on image understanding tasks with fewer parameters and less computation. This makes it a promising alternative to traditional CNNbased approaches, especially in scenarios where computational resources are limited. The formula for computing the attention weights in the multi-head self-attention mechanism is given by:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})$$

where Q, K, and V are the query, key, and value matrices, respectively, and d_k is the dimension of the key vectors.

The feedforward network (FFN) is a sublayer of the transformer block that applies a non-linear transformation to the input vector. The formula for computing the output of the FFN is given by: $FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2$ where x is the input vector, W_1, b_1, W_2, and b_2 are learnable weight matrices and bias vectors, respectively. The FFN is used to introduce non-linearity into the transformer block and increase its expressive power. The layer normalization and residual connection are also sublayers of the transformer block. The formula for computing the output of the layer normalization and residual connection is given by: $LayerNorm(x + Sublayer(x))$ where $Sublayer(x)$ is the output of the self-attention or feed-forward sublayer, and $LayerNorm$ is a learnable normalization function. The layer normalization and residual connection are used to improve the stability and convergence of the transformer block during training.

In general, the feed-forward network and layer normalization are used in the transformer block to introduce non-linearity and improve the stability and convergence of the model during training. These sublayers are applied after the self-attention sublayer, which is responsible for capturing the dependencies between the input tokens. DeiT has the same architecture as ViT, except for the input token part, which includes an additional distillation token.The primary reason for choosing DeiT over the vanilla vision transformer is that ViT does not perform well when trained on limited data. However, DeiT overcomes this limitation through distillation. Since our dataset is not extensive, we preferred to choose DeiT, which is also computationally less demanding.

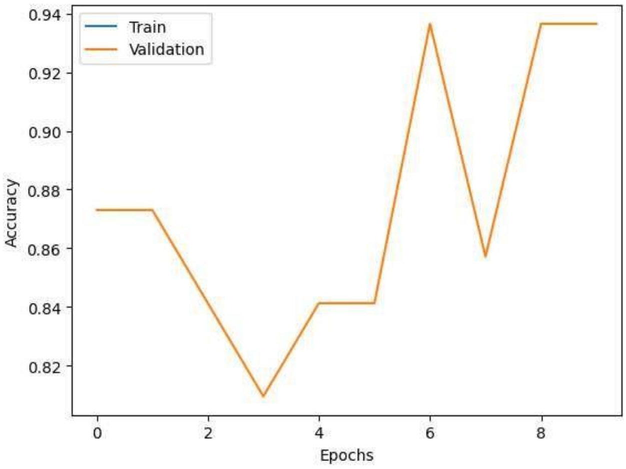

Fig. 5. Validation curve of DeiT's model

Architecture: The DeiT model was trained for a total of 9 epochs, achieving a validation accuracy of 93.65% and a validation loss of 0.321. During both the training and validation phases, the LabelSmoothingCrossEntropy loss function was utilized.

$$\text{LabelSmoothingCrossEntropy} = -\sum_{i=1}^{N}(q_i \cdot \log(p_i))\ (2)$$

In this formula: In this formula:

1) We take the sum $\sum_{i=1}^{N}$ over all classes (from 1 to N).
2) For each class, we multiply the smoothed ground truth probability q_i by the logarithm of the predicted probability p_i.
3) Finally, we negate the result to obtain the loss value.

IV. RESULTS

The Data-Efficient Image Transformers (DeiT) model, with fine-tuning for ILD detection, achieved an impressive accuracy of 91%. In comparison, ResNet50 and VGG16 yielded accuracies of 87% and 56%, respectively. These results underscore the efficacy of DeiT as a powerful tool for ILD classification, particularly in cases where data availability is limited. The success of the DeiT model in achieving an accuracy of 91% in ILD classification is a result of several key factors. DeiT, a novel convolution-free neural network architecture, represents a significant departure from traditional convolutional neural networks (CNNs) used in image analysis. Instead of relying on convolutional layers to extract features, DeiT employs selfattention mechanisms, such as multi-head self-attention, to capture complex spatial relationships within the input images. This innovative approach allows the model to attend to different parts of the image, irrespective of the size of the convolutional kernel. As a result, DeiT can recognize intricate patterns and dependencies that may be vital for accurate ILD detection.

Additionally, the DeiT model's use of self-attention mechanisms provides an edge in capturing long-range dependencies between different parts of the image, enabling it to learn more abstract and complex representations of the input image. This capability is particularly crucial in the context of ILD detection, as it ensures the model can recognize subtle abnormalities within the interstitial tissue, even in the presence of variations in image quality and contrast.

Furthermore, DeiT's approach to handling input images, where they are divided into non-overlapping patches and processed as a sequence of tokens, allows for more flexible and adaptable image understanding. Each token is associated with a learnable embedding vector, facilitating the model's adaptability to varying imaging conditions, including changes in orientation, illumination, and color balance. In essence, DeiT's structure equips it to be invariant to variations in image color and brightness, which can be particularly challenging in the analysis of medical images.

Another crucial aspect that contributed to DeiT's outstanding performance is the fine-tuning process. By keeping the foundational layers of the model frozen and adding specific layers like dropout and a linear layer for ILD classification, the model was fine-tuned to excel in the context of this specific task. This fine-tuning process enabled DeiT to focus on the intricacies of ILD detection, making it more adept at recognizing the telltale signs of the disease in medical images.

In essence, DeiT's revolutionary architecture, combined with its capacity to capture long-range dependencies, adapt to variations in imaging conditions, and its fine-tune for ILD classification, made it the top performer in our study. These results emphasize the potential of innovative neural network architectures and the importance of tailoring models to the specific requirements of medical image analysis, shedding light on the path to more accurate and reliable ILD detection in clinical practice.

The research dataset utilized in this study comprises a collection of CT scans of lungs in DICOM format from individuals diagnosed with Interstitial Lung Disease (ILD). The dataset was divided into three specific ILD subtypes, including Interstitial Pneumonia, Adenocarcinoma, and Scleroderma, and a control group of lung images from individuals without ILD. Our preprocessing techniques included Min-Max Scaling Intensity Normalization to standardize pixel intensities and Histogram Equalization for enhanced image contrast. Additionally, resizing and cropping were applied to optimize the images for ILD detection. Data augmentation techniques such as horizontal and vertical flipping, color jittering, and random erasing were employed to enhance the diversity of training samples. These preprocessing methods significantly improved the quality and interpretability of the images.

For classification, we divided the preprocessed data into training, validation, and testing sets. We employed various deep learning and transfer learning techniques. Notably, the Data-Efficient Image Transformers (DeiT) model, with finetuning for ILD detection, achieved an impressive accuracy of 91%. In comparison, ResNet50 and VGG16 yielded accuracies of 87% and 56%, respectively. These results underscore the efficacy of DeiT as a powerful tool for ILD classification, particularly in cases where data availability is limited.

V. CONCLUSION

In conclusion, our research has made significant strides in the realm of Interstitial Lung Disease (ILD) detection through the utilization of deep learning and transfer learning techniques. While traditional deep learning models such as ResNet50 and VGG16 demonstrated commendable performance with accuracies of 87% and 56%, respectively, the real breakthrough came with the adoption of the Data-Efficient Image Transformers (DeiT) model, which delivered an outstanding accuracy of 91%.

These results underscore the pivotal role of innovative neural network architectures in the realm of medical image analysis. The DeiT model's distinct approach, centered around self-attention mechanisms, allows it to capture intricate spatial relationships and long-range dependencies within images, making it exceptionally well-suited for detecting subtle abnormalities in the interstitial tissue, even in the face of variations in image quality, contrast, and orientation.

As we look ahead, one promising avenue for further enhancing ILD detection accuracy is the employment of HighResolution CT (HRCT) scans. While our research predominantly utilized standard CT scans, the incorporation of HRCT imaging promises to unlock even more intricate details. These enhanced details have the potential further to boost the accuracy of the meticulous DeiT model. The intricate and nuanced characteristics of HRCT images align well with the capabilities of DeiT, offering the prospect of achieving even greater accuracy in ILD detection. Future research endeavors may benefit significantly from the utilization of HRCT, propelling the field of ILD diagnosis and potentially revolutionizing the accuracy and reliability of this critical clinical practice.

REFERENCES

[1] A. V. Samarelli, R. Tonelli, A. Marchioni, G. Bruzzi, F. Gozzi, D. Andrisani, I. Castaniere, L. Manicardi, A. Moretti, L. Tabb`ı, *et al.*, "Fibrotic idiopathic interstitial lung disease: The molecular and cellular key players," *International Journal of Molecular Sciences*, vol. 22, no. 16, p. 8952, 2021.

[2] Y. Kondoh, S. Makino, T. Ogura, T. Suda, H. Tomioka, H. Amano, M. Anraku, N. Enomoto, T. Fujii, T. Fujisawa, *et al.*, "2020 guide for the diagnosis and treatment of interstitial lung disease associated with connective tissue disease," *Respiratory Investigation*, vol. 59, no. 6, pp. 709–740, 2021.

[3] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in vision: A survey," *ACM computing surveys (CSUR)*, vol. 54, no. 10s, pp. 1–41, 2022.

[4] S. Iqbal, A. N. Qureshi, J. Li, and T. Mahmood, "On the analyses of medical images using traditional machine learning techniques and convolutional neural networks," *Archives of Computational Methods in Engineering*, vol. 30, no. 5, pp. 3173–3233, 2023.

[5] X. Lu, W. Gong, Z. Peng, F. Zeng, and F. Liu, "High resolution ct imaging dynamic follow-up study of novel coronavirus pneumonia," *Frontiers in Medicine*, vol. 7, p. 168, 2020.

[6] Denis Hadjiliadis, Paul F. Harron, "All lending club loan data. kaggle." https://www.pennmedicine.org/for-patients-and-visitors/ patient-information/conditions-treated-a-to-z/lung-disease#:~: text=Definition,and%20out%20of%20the%20 lungs., 2022.

[7] H. M. Shyni and E. Chitra, "A comparative study of x-ray and ct images in covid-19 detection using image processing and deep learning techniques," *Computer Methods and Programs in Biomedicine Update*, vol. 2, p. 100054, 2022.

[8] S. S. Bamber and T. Vishvakarma, "Medical image classification for alzheimer's using a deep learning approach," *Journal of Engineering and Applied Science*, vol. 70, no. 1, p. 54, 2023.

[9] J. G. Park and C. Lee, "Skull stripping based on region growing for magnetic resonance brain images," *NeuroImage*, vol. 47, no. 4, pp. 1394– 1407, 2009.

[10] S. Hossain, A. Chakrabarty, T. R. Gadekallu, M. Alazab, and M. J. Piran, "Vision transformers, ensemble model, and transfer learning leveraging explainable ai for brain tumor detection and classification," *IEEE Journal of Biomedical and Health Informatics*, 2023.

[11] F. J. M. Shamrat, S. Azam, A. Karim, K. Ahmed, F. M. Bui, and F. De Boer, "High-precision multiclass classification of lung disease through customized mobilenetv2 from chest x-ray images," *Computers in Biology and Medicine*, vol. 155, p. 106646, 2023.

[12] J. Arslan, G. Samarasinghe, A. Sowmya, K. Benke, L. Hodgson, R. Guymer, and P. Baird, "Deep learning applied to automated segmentation of geographic atrophy in fundus autofluorescence images," *Translational vision science technology*, vol. 10, 07 2021.

[13] J. Morgado, T. Pereira, F. Silva, C. Freitas, E. Negrao, B. F. de Lima,˜ M. C. da Silva, A. J. Madureira, I. Ramos, V. Hespanhol, *et al.*, "Machine learning and feature selection methods for egfr mutation status prediction in lung cancer," *Applied Sciences*, vol. 11, no. 7, p. 3273.

[14] R. Clay, B. Bartholmai, B. Zhou, R. Karwoski, T. Peikert, T.

Osborn, S. Rajagopalan, S. Kalra, and X. Zhang, "Assessment of interstitial lung disease using lung ultrasound surface wave elastography–a novel technique with clinicoradiologic correlates," *Journal of thoracic imaging*, vol. 34, no. 5, p. 313, 2019.

[15] S. Bharati, P. Podder, and M. R. H. Mondal, "Hybrid deep learning for detecting lung diseases from x-ray images," *Informatics in Medicine Unlocked*, vol. 20, p. 100391, 2020.

[16] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[17] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, *et al.*, "A survey on vision transformer," *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 1, pp. 87–110, 2022.

[18] W. D. Bidgood Jr, S. C. Horii, F. W. Prior, and D. E. Van Syckle, "Understanding and using dicom, the data interchange standard for biomedical imaging," *Journal of the American Medical Informatics Association*, vol. 4, no. 3, pp. 199–212, 1997.

[19] M. Kociolek, M. Strzelecki, and S. Szymajda, "On the influence of the image normalization scheme on texture classification accuracy," in *2018 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, pp. 152–157, 2018.

[20] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jegou, "Training data-efficient image transformers & distillation´ through attention," in *International conference on machine learning*, pp. 10347–10357, PMLR, 2021.