Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 7, No. 4, October 2025, pp: 1158-1177. e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

COV-TViT: An Improved Diagnostic System for COVID Pneumonitis Utilizing Transfer Learning and Vision Transformer on X-Ray Images

Sunil Kumar[1], Amar Pal Yadav[2], Neha Nandal[3], Vishal Awasthi[1], Luxmi Sapra[4], Prachi Chhabra[5],

[1] School of Engineering and Technology (UIET), CSJM University, Kanpur, India.
[2] Department of Computer Science and Engineering, Noida Institute of Engineering and Technology, Greater Noida, India.
[3] Department of Computer Science and Engineering, Geethanjali College of Engineering and Technology, Hyderabad, India.
[4] School of Computing, Graphic Era Hill University, Dehradun, India.
[5] Department of Information Technology, JSS Academy of Technical Education Noida, Noida, India.

Corresponding author: Sunil Kumar (e-mail: sunilymca24@gmail.com) , **Author(s) Email**: Amar Pal Yadav (e-mail: challenge_amar@rediffmail.com), Neha Nandal (e-mail:neha28nandal@gmail.com), Vishal Awasthi (e-mail: vishalawasthi@csjmu.ac.in), Luxmi Sapra (e-mail: luxmi.sapra@gmail.com), Prachi Chhabra (e-mail: prachichhabra@jssaten.ac.in)

Abstract **COVID is a contagious lung ailment that continues to be a world curse, and it remains a highly infectious respiratory disease with global health implications. Traditional diagnostic methods, such as RT-PCR, though widely used, are often constrained by high costs, limited accessibility, and delayed results. In contrast, radiology for lung disease detection has been proven advantageous for identifying deformities, and chest X-rays are the most preferred radiological method due to their non-invasive nature. To address these limitations, this study aims to develop an efficient, automated diagnostic system leveraging radiological imaging, specifically X-rays, which are cost-effective and widely available. The primary contribution of this research is the introduction of COV-TViT, a novel deep learning framework that integrates transfer learning with Vision Transformer (ViT) architecture for the accurate detection of COVID pneumonitis. The proposed method is evaluated using the COVID-QU-Ex dataset, which comprises a balanced set of X-ray images from COVID positive and healthy individuals. Methodologically, the system employs pre-trained convolutional neural networks (CNNs), specifically VGG16 and VGG19 (Visual Geometry Group), for transfer learning, followed by fine tuning to enhance feature extraction. The ViT model, known for its self-attention mechanism, is then applied to capture complex spatial dependencies in the X-ray images, enabling robust classification. Experimental results demonstrate that COV-TViT achieves a classification accuracy of 98.96% and an F1 score of 96.21%, outperforming traditional CNN based transfer learning models in several scenarios. These findings underscore the model's potential for high-precision COVID pneumonitis detection. The proposed approach significantly transforms classification tasks using self-attention mechanisms to extract features and learn representations. Overall, the proposed diagnostic system COV-TViT can be advantageous in the fundamental identification of COVID pneumonitis.**

Keywords **Convolution Neural Network, Deep Learning, Machine Learning, Self-Attention, VGG, ViT.**

I. Introduction

In March 2020, the World Health Organization (WHO) stated that COVID-19, a form of pneumonitis as an eruption, had become an epidemic, starting in Wuhan, China, and then expanding globally [1]. COVID pneumonitis is tested using two common tests: an antibody test and a viral test [2]. Antibody tests detect antibodies in blood samples to find out if the patient has previously been exposed to the COVID pneumonitis virus [3]. Viral tests can diagnose an ongoing infection, starting with antigen tests and nucleic acid amplification tests (NAATs). The virus that originates COVID-19, SARS-CoV-2, is found by viral testing and the collection of a sample from the nose or mouth of the

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1037

infected person. People with COVID pneumonitis symptoms or who have had direct contact with someone who has tested positive can be recommended to have a viral test. NAATs detect the genetic material of the virus in samples taken from the respiratory tract. They work by amplifying nucleic acids and detecting the ribonucleic acid (RNA) sequences that encompass the genetic material of the SARS-CoV-2 virus. There are two methods of NAATs: isothermal amplification and reverse transcription-polymerase chain reaction (RT-PCR) tests [4]. Antigen tests are immunoassays that detect viral antigens to identify current viral infections. They are inexpensive and can provide results in approximately 15 minutes [5]. Conventionally, the most often used NAAT test is RT-PCR. Despite its specificity, it is complex and time-consuming, as it can take up to two days to produce results. NAAT is expensive and less accessible compared to radiology, which is readily available in almost every health clinic [6].

In the latest studies, beyond direct viral detection, recent research has explored radiological imaging as a diagnostic aid for COVID pneumonitis. Chest radiography seems promising for detecting traces of pneumonia caused by COVID. Chest X-ray (CXR) imaging has shown considerable promise in identifying characteristic COVID-19 pneumonia patterns associated with the symptoms. CXR scans are available almost instantly after scanning, and the approach is an established method with numerous papers published by researchers [7-9]. Computed tomography (CT) scans are costlier than X-ray scans, rendering them unaffordable for specific individuals. CT scan technology is less common in underdeveloped nations than X-ray machines [7]. Therefore, there is a need to develop reliable systems for automated detection that will help ease the burden on the healthcare system with X-rays. CXR offers distinct advantages: it is readily available, provides near instant results post scan, is relatively inexpensive, and leverages a well-established clinical infrastructure with extensive prior research. CT scans are pricier and less available than CXR scans, making CXR a more practical choice for widespread screening or triage [8]. Consequently, there is growing interest in leveraging CXR for COVID-19 detection.

With the assistance of machine learning (ML) in healthcare, which uses radiological scans for the detection of COVID-19, this can be an alternate option [10]. The study employed transfer learning to utilize a pre trained Convolutional Neural Network (CNN) model and ViT methods [11] to classify X-ray images from the publicly accessible COVID-QU-Ex dataset [12] for research purposes. This capability enables the model to identify subtle and complex patterns indicative of COVID-19 pneumonia, potentially capturing features less discernible to CNNs. The system undergoes training and evaluation using the publicly available COVID-QU-Ex dataset.

Although CXR presents a viable and accessible imaging modality for detecting COVID-19-related pneumonia, manual interpretation by radiologists is time consuming and subject to variability and can overwhelm healthcare systems during high volume pandemic periods. While ML, particularly deep learning applied to radiological scans, has emerged as a potential solution for automated COVID-19 detection [10], the existing approaches often rely heavily on standard CNNs or their transfer learned variants. A significant gap exists in thoroughly exploring and optimizing the application of the novel ViT architecture [11], originally designed for natural images, specifically for the task of COVID-19 diagnosis from CXR images. There is a need for robust, automated systems that leverage cutting edge deep learning paradigms like ViT to maximize efficiency in interpreting readily available CXR scans for COVID-19.

This research presents COV-TViT, a distinctive diagnostic system designed for COVID-19 detection through transfer learning and a ViT based approach. Our research highlights the efficacy of transfer learning and ViT in effectively meeting the pressing need for prompt and accurate COVID-19 diagnosis using non-invasive imaging techniques. This method exemplifies the necessity of employing transfer learning procedures and provides valuable insights to researchers examining COVID-19 patients through X-ray analysis. In the context of transfer learning, our strategy involves using fine-tuned preexisting VGG16 and VGG19 models [13]. It allows us to use the previously learned information and hierarchical representations. It enables the networks to efficiently capture unique patterns associated with COVID-19 in X-ray images, enabling precise and effective identification of the illness. The methodology includes the utilization of pre-trained VGG16 and VGG19 CNNs in conjunction with a customized ViT for transfer learning. Initially, fine-tuned VGG16 and VGG19 models were used to leverage their learned features from extensive image datasets, effectively capturing unique COVID-19 patterns in CXR images. Secondly, and more innovatively, the ViT architecture [11] is adapted for the task. ViT breaks down the CXR image into patches and uses self-attention mechanisms to understand the image's long range connections and spatial relationships. The ViT, initially designed for image processing, has been adapted to accurately identify intricate patterns and subtle indicators of COVID-19 in X-ray images. The ViT can distinguish specific radiological attributes related to the virus using self-attention and hierarchical feature extraction techniques [14]. This capability allows the ViT to achieve high accuracy and efficiency in diagnosing

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v7i4.1037

viruses. The implementation of ViT in this context demonstrates substantial promise in improving the prompt identification and surveillance of COVID-19. The main objective of this research is to create and validate the COV-TViT framework, an automated diagnostic system aimed at promptly and accurately detecting COVID-19 infections through the analysis of accessible CXR images. This goal is accomplished by synergistically combining the strengths of transfer learning from established CNN architectures with the advanced pattern recognition capabilities of the ViT paradigm.

While RT-PCR remains the diagnostic gold standard for COVID-19, its limited sensitivity during early infection and delayed turnaround time hinder timely intervention. Radiological imaging, particularly X-rays, offers rapid and accessible screening but suffers from interpretive variability and overlap with other pneumonias [15]. The existing AI-based methods often rely on small, imbalanced datasets and conventional CNNs with limited receptive fields, restricting their ability to generalize and capture global contextual features. Although ViTs offer enhanced spatial modeling via self-attention, their application to medical imaging is constrained by high computational demands, lack of locality bias, and limited optimization for CXR specific patterns.

These comprehensive limitations create a significant knowledge gap in COVID-19 diagnostic capabilities. Current approaches fail to effectively combine the accessibility of X-ray imaging with the pattern recognition capabilities of advanced deep learning architectures. Specifically, there is insufficient research on optimizing ViT architectures for COVID-19 detection, particularly regarding transfer learning strategies that can address data scarcity while maintaining diagnostic effectiveness. The need for robust, automated systems that can provide rapid, accurate COVID-19 detection from readily available imaging modalities remains unmet.

Additionally, early transfer learning pipelines employing DenseNet-121, SqueezeNet, ResNet50, and ResNet18 on the COVID-Xray-5k dataset achieved accuracies up to 94.1% but often overfitted small sample sizes and remained constrained by local receptive fields [8]. COVIDNet's Projection Expansion Projection Extension (PEPX) architecture further improved average accuracy to 94.10% by pretraining on extensive public cohorts and embedding extracted features into a visual transformer [9] [10]. Hybrid ViT-based solutions, such as VitCNX (98.21% accuracy, 99.91% AUPR) [11] and DAViT (97% F1 score, 96% AUC), demonstrate superior global context modeling. However, they require large, balanced datasets, and substantial computational resources. Comparative analyses have identified ResNet and VGG19 as the leading CNN backbones in COVID-19 CXR classification [16], while U-Net-based segmentation [17] and Grad-CAM-guided interpretability [14] have been proposed to bolster clinical applicability. These studies underscore a persistent tradeoff between local feature extraction, global dependency modeling, data efficiency, and scalability. In this context, we proposed COV-TViT, which synergistically combines fine-tuned VGG16/19 CNNs for robust local feature learning with a customized ViT for enhanced global contextual understanding, aiming to deliver accurate, efficient, and scalable COVID-19 detection from X-ray images.

The proposed COV-TViT diagnostic framework integrates transfer learning and ViT architectures to enable accurate and efficient COVID-19 detection, demonstrating a significant advancement in ML assisted medical imaging:

1. Novel Framework (COV-TViT): Our innovative COV-TViT diagnostic framework is tailored for detecting COVID pneumonitis from CXR scans. This framework uniquely combines transfer learning on a pre trained CNN (VGG16 and VGG19) with a tailored implementation of the ViT architecture, offering a new approach to leveraging both established and emerging deep learning techniques for this critical task.
2. Demonstrated Efficacy of Transfer Learning: This study shows that adjusting the VGG16 and VGG19 models [13] through transfer learning is highly effective in identifying important features related to COVID-19 pneumonia in CXR images. It validates the utility of leveraging pre-existing hierarchical representations for efficient and accurate medical image analysis in this domain.
3. Pioneering ViT Adaptation for CXR: We introduce a groundbreaking adaptation and thorough evaluation of the ViT model [11] for diagnosing COVID-19 using CXR scans. The work showcases the model's capability, via self-attention mechanisms [14], to capture complex spatial relationships and subtle pathological patterns indicative of COVID-19, potentially surpassing the limitations of local receptive fields in traditional CNNs for this application.
4. Practical Diagnostic Advancement: The COV-TViT system's high accuracy in automated COVID-19 detection from widely available CXR scans offers a practical solution to ease the strain on healthcare systems. It offers a pathway toward faster triage, reduced reliance on slower traditional tests like RT-PCR in certain scenarios, and improved resource management during pandemic surges.

The article's structure is as follows: Section 2 delves into the essential essence of the inquiry,

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1037

considering contemporary academic research. Section 3 thoroughly expounds on the materials and methodologies used in the study. Section 4 offers an extensive examination and elucidation of the study results, improving the investigation outcomes' exposition. Section 5 discusses the research and comparative analysis. The investigation is concluded within Section 6.

II. Related Work

Chest radiography has been discovered to help detect lung abnormalities and diagnose lung disorders. At the onset of the pandemic, the number of samples of chest X-rays was too low to develop generic deep learning methods. However, over the years, researchers have accumulated sufficient data to make general models that can be employed to assist clinicians. State of the art brought forth the point that conventional ML approaches, transfer learning based methods, and ViT-based methods were used to accomplish this task.

S. Minaee et al. created the COVID-Xray-5k dataset, which contains around 5000 chest X-ray scanned images. A professional radiologist labels the X-ray scans for COVID-19 classification. The employed dataset enabled the DeepCovid model to trace COVID-19. They used transfer learning for training four well known CNNs: DenseNet-121, SqueezeNet, ResNet50, and ResNet18 [8]. L. Wang et al. [9] presented COVIDNet, a PEPX architecture, to classify infections as normal, COVID pneumonitis, and non COVID. The backbone network underwent training utilizing extensive publicly available datasets to identify and capture aberrant features in COVID-19 diagnoses, such as consolidation, ground glass opacity (GGO), and others. Subsequently, the embedded features derived from the underlying network were employed as a corpus to train the visual transformer. The experimental findings yielded the maximum average accuracy of 94.10% on one of the employed test datasets. The suggested model has been verified to attain state of the art performance in diagnosing COVID-19 [10]. The researcher demonstrated VitCNX, an advanced deep learning solution for COVID pneumonitis image identification that utilizes ViT and ConvNeXt. The recommended VitCNX model contrasted against prominent models built with deep learning, such as E-NetV2, ResNet-50, DenseNet, Swin Transformer, ViT, and ConvNeXt. With a recall of 99.07%, accuracy of 98.21%, F1 score of 98.55%, AUC of 99.85%, and AUPR of 99.91%, VitCNX outperformed. VitCNX achieved outstanding results in the three classification task, with a precision of 96.68%, accuracy of 96.96%, and F1 score of 96.31%. These results highlight the outstanding picture classification capabilities of VitCNX. The proposed VitCNX model intends to help identify COVID-19 patients [11]. A deep learning pipeline using ViT was employed to identify COVID pneumonitis from the X-rays. The researchers collected a total of 30,000 images of X-rays from three publicly available datasets. The proposed transformer model demonstrated extraordinary accuracy in differentiating COVID from normal X-rays, achieving a 98% accuracy rate and a 99% AUC score in the binary classification process. The multi class classification test yielded a 92% accuracy and a 98% AUC score for distinguishing X-rays in patients. The test dataset was assessed using commonly used models, including EfficientNetB0, InceptionV3, ResNet50, MobileNetV3, Xception, and DenseNet-121, which served as the reference models. The transformer model demonstrated superior performance across all criteria. Grad-CAM visualization was applied to enhance the approach's comprehensibility [14]. Hussain et al. [15] conducted a series of classification experiments including several classes, such as two classes for normal and COVID-19, three classes for normal, COVID-19, and pneumonia bacteria, and four classes for normal, COVID-19, pneumonia bacteria, and pneumonia viral. “CoroDet” is the suggested technique, which comprises a novel 22 layer CNN model. In their study, A. Narin et al. [16] examined the effectiveness of a deep transfer learning method that used five distinct CNN models for three binary categories. Transfer learning offers a significant advantage in data training by allowing the use of a smaller amount of data. Among all the models trained in the research, ResNet achieved the best accuracy. The COVQU dataset included 18,479 CXRs of subjects with normal lung problems, COVID cases, and lung capacity anomalies unrelated to COVID. A modified version of the U-Net network was presented by the researchers for the purpose of lung segmentation and classification. This network makes use of seven sophisticated CNN models, one of which is the ChexNet model that was suggested [17]. In their study, Iqbal et al. [18] conducted four distinct classifications of classes, including Normal, COVID-19, Pneumonia Bacterial, and Viral. In addition, they successfully conducted three class classifications: normal, COVID-19, and pneumonia. The classifications were successfully conducted on a range of carefully prepared datasets using the CoroNet model introduced by the authors. The Xception CNN architecture served as the foundation for the proposed model. The Xception architecture is a 71-layer version of the Inception architecture. The proposed domain adapted vision transformer (DAViT) is a hybrid vision transformer CNN model with domain adaptation that fuses global and local features. DAViT achieved pneumonia detection with a 97% F1 score and 96% AUC, surpassing twelve baseline methods [19]. M. Rahaman et al. used deep transfer learning to analyze 15 pre trained CNN models and found that VGG19 performed well with an accuracy of 89.30% [20]. They

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1037

identified COVID using data from two publicly accessible databases: "COVID-19 image data collection" [21] and "Chest X-ray images (Pneumonia)" [22]. The proposed work created a trustworthy deep learning model for reliably classifying COVID-19 X-rays. The suggested technique extracts deeper features from images using an ensemble method and then employs global second order pooling to obtain higher global image features. In addition, images are segmented into patches and positions embedded before being examined independently using a ViT approach. The Covid ChestX-ray-15k dataset achieved 97.84% accuracy, 96.76% sensitivity, and 96.80% precision [23].

III. Material and Methods

A. Dataset

The publicly accessible COVID-QU-Ex dataset was employed as the primary source for experimental evaluation in this study. Originally, the COVID-QU-Ex dataset consisted of 33,920 curated chest X-ray images categorized into three diagnostic classes: COVID (11,956), normal (10,701), and non COVID infections (11,263). The distinguishing feature of the dataset is the inclusion of accurate lung segmentation masks for all images, facilitating detailed anatomical feature analysis [12].

B. Optimized Preprocessing Framework

The preprocessing methodology employed a targeted stage and processing based approach leveraging the COVID-QU-Ex dataset's unique segmentation masks to optimize preprocessed image. The stages are:

1. Anatomical Region Isolation: Binary lung masks were applied to isolate pulmonary structures, suppressing non diagnostic regions (mediastinum, thoracic cage) by nullifying pixel intensities outside segmented areas. This focused computational resources on pathologically relevant zones [12].

2. Adaptive Contrast Enhancement: Contrast Limited Adaptive Histogram Equalization (CLAHE) with a clip limit of 2.0 and an 8×8 tile grid was applied exclusively within lung boundaries. This amplified subtle radiographic patterns (e.g., GGO) while preventing noise amplification in homogeneous tissue regions [24].

3. Mask-Constrained Normalization: z-score normalization also known as standard scaling, utilized intensity statistics (mean μ, standard deviation σ) calculated solely within lung regions as illustrated through Eq. (1) [25].

$$z = \frac{(x - \mu)}{s} \quad (1)$$

where, z = Output/transformed data, x = Input data. Standard scaling was applied during preprocessing to enhance training efficiency and model stability by ensuring all features have zero mean and unit variance, thus enabling equal treatment across scales.

4. Multimodal Input Construction: The input comprised three diagnostically complementary representations: normalized grayscale images, binary lung masks, and CLAHE enhanced images. These were resampled to a resolution of 224×224 and stacked as input channels to enrich feature diversity and spatial context [25].

5. Augmentation Strategy: The augmentation strategy, applied exclusively to the training data, included mask preserving horizontal flipping (p = 0.5) and lung centric rotation (±10° around the anatomical centroid to enhance model generalization) [9]. The preprocessing and employed model integration workflow is presented in Fig. 1:

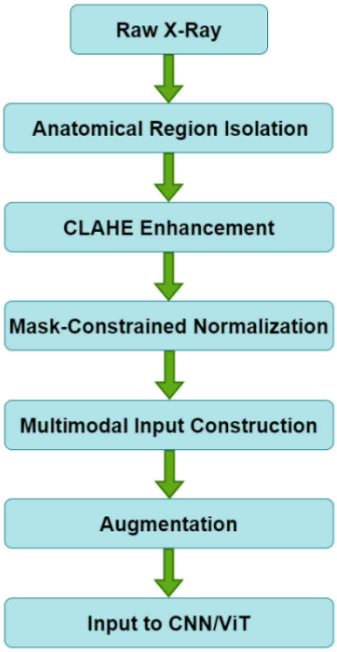

Raw X-Ray
Anatomical Region Isolation
CLAHE Enhancement
Mask-Constrained Normalization
Multimodal Input Construction
Augmentation
Input to CNN/ViT

Fig. 1. Preprocessing Steps

C. Methodology

This work discusses the operational principles of the ViT when it was applied to images. The critical steps of the ViT's operation are outlined, including input encoding, positional encoding, self-attention mechanism, ViT encoder, classification component, and training and fine tuning operations on preprocessed X-ray images. The incorporation of the suggested diagnostic system is a crucial aspect. Fig. 2 illustrates the framework of the proposed COV-TViT Diagnostic System, which utilizes the employed dataset for training. The framework consists of many steps: preparatory processing, dataset partitioning, applying transfer learning techniques utilizing the VGG-16 model, and integrating the ViT structure.

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1037

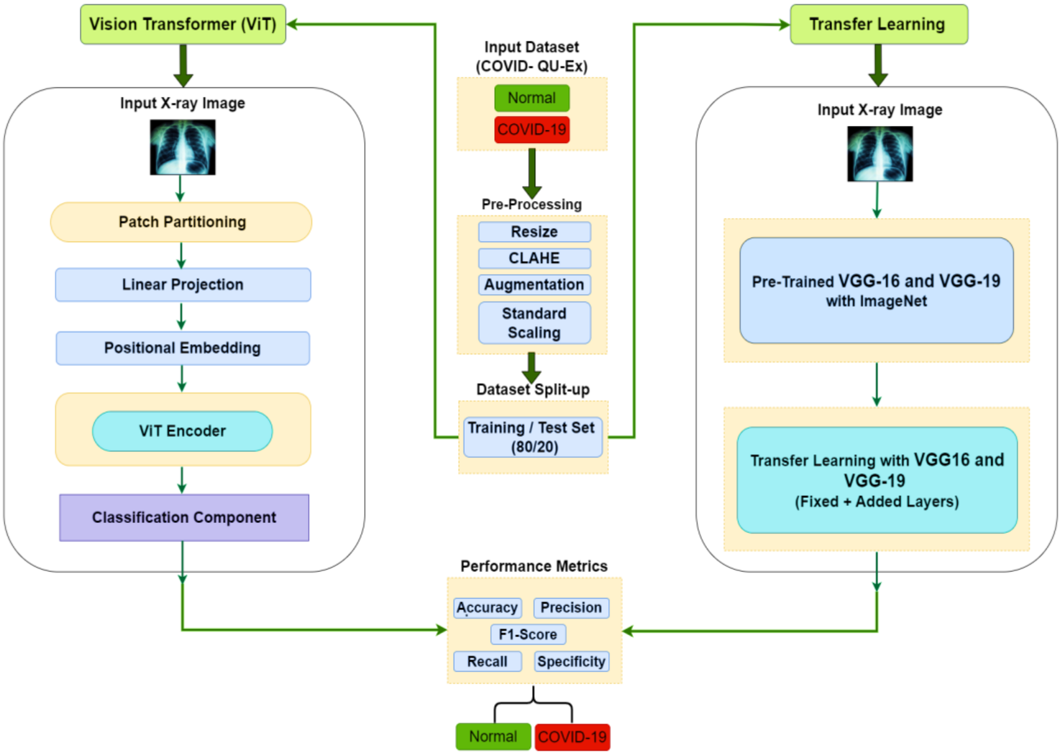


Fig. 2. The COV-TViT Diagnostic System Framework

D. Convolution Neural Network

CNN extracts high level features from the input image using convolution operation. A CNN can extract shift invariant feature maps by applying the shared weight architecture of filters or convolution kernels [23]. A CNN is a distinct neural network that incorporates convolution operations inside one or more hidden layers. These convolution operations produce a feature map from the input matrix of the respective layer, which then serves as the input for the subsequent layer. Eq. (2) presented the CNN concept [26, 27]. Pooling layers are often included next to convolution layers to decrease the dimensionality of the data. Additionally, fully linked layers, analogous to conventional multilayer perceptrons, are commonly formed [9].

$$S(i,j) = (I * K)(i,j) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} I(i+m, j+n) \cdot K(m,n) \quad (2)$$

where, $S(i,j)$: output feature map at position (i,j), $I(i,j)$: input pixel value at position (i,j), K(m,n): filter/kernel weight at position (m,n).

E. VGG16 and VGG19

The VGG16 and VGG19 models are renowned for their straightforward and uniform architecture. VGG16 and VGG19 consist of 16 and 19 convolutional layers that use tiny 3x3 filters and pooling layers, respectively, followed by three fully connected layers, with VGG19 extending VGG16's architecture via additional convolutional blocks. Owing to their stable performance and standardized design, both were widely adopted as benchmark CNN models presented through Eq. (3) [13, 25].

$$X^{(l)} = f\left(Conv\left(X^{(l-1)}, W^{(l)}\right) + b^{(l)}\right) \quad (3)$$

For each convolutional layer l, given input feature map $X^{(l-1)}$, filter weights $W^{(l)}$, biases $b^{(l)}$, the layer output is $X^{(l)}$.

These models demonstrated strong efficacy in several COVID pneumonitis X-ray classification tasks and yielded promising accuracy results. Nevertheless, the VGG16 and VGG19 models unveiled several parameters, resulting in a significant computing burden.

F. Transfer Learning

Transfer learning is a strategic approach that aims to overcome the constraints of current CNN designs by using pre-trained models and adapting them to novel but interconnected tasks. The process entails using early layers to extract features and fine tuning the top fully linked layers presented through Eq. (4) [8]. This equation provides a foundation for adapting a pre-trained CNN to a new task by adjusting the input data and reusing the learned mapping for prediction.

$$f_{int} := f_S^* \circ T_X \in \{f_{int} \mid f_{int}: X_T \to Y_S\} \quad (4)$$

where, f_S^* is the pre-trained model, T_X is the input transport mapping function that transforms inputs from

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1037

the target domain input space X_T to the source domain input space X_S. X_T is the input space of the target domain. Y_S is the output/label space of the source domain corresponding to f_S^*, f_{int} is the composed intermediate function, created by applying the input transport mapping T_X followed by the pretrained model.

This technique's use of preexisting models allows for transferring features, representations, and weights from one job to another. This process may accelerate the training process and improve accuracy, especially in scenarios where training data is scarce. Nevertheless, the method's efficacy relies on the degree of similarity between the original and desired tasks, since substantial differences in domain tend to result in subpar performance. The meticulous adjustment of hyper-parameters is crucial to mitigate the risk of overfitting, while carefully selecting a suitable pre-trained model is pivotal in attaining the intended outputs [20, 26]. The approach included using an established methodology to address the challenges posed by the COVID-19 pandemic. Transfer learning must use large datasets, including millions of data points, for practical training. The implementation improved operational efficiency and greater resource allocation while training new models. Transfer learning proves to be of great use in datasets with insufficient labeling [7, 18]. Using pre-trained weights and acquired features from the COVID-QU-Ex dataset substantially improves the model's performance.

Two cases were used to show how VGG16 can be used with transfer learning. In the first case, the VGG16 model's original upper layers were kept, and the weights that had already been trained on a different task were left alone [28]. In the second case, new layers were added on top of the original highest levels of the VGG16 model. There were three tightly linked layers in the model design. They were 256, 256, and 128, respectively. A softmax activation function and two units for binary classification came after these layers. The training focused only on updating the output layer weights, enabling the VGG16 model to refine its acquired features. A demonstration was carried out using the VGG16 model presented to better understand the use of CNN architectures based on transfer learning (refer to Fig. 3). The training procedure consisted of training each model for 50 and 100 epochs, with a batch size of 64.

In this investigation, the loss function was binary cross entropy. The Adam optimizer, alongside a learning rate of 0.002, was utilized. A total of eight models were trained using this approach. Using the Adam optimizer enhances the likelihood of reducing the error rate to its maximum extent. The primary function of the loss of binary cross entropy throughout learning is to compute and reduce errors. Table 1 displays the hyper-parameters used in the research.

Table 1. **Hyper-Parameters**

Hyper-parameter	Instance
Optimizer	Adam
Learning Rate	0.002
Loss Function	Binary Cross Entropy
Batch Size	64
Epochs	50 and 100

G. Transfer Learning Configuration

In our transfer learning setup, VGG16 and VGG19 backbones pre-trained on ImageNet were adapted to the COVID-QU-Ex dataset by freezing the first three convolutional blocks (10 layers in VGG16; 12 in VGG19), which preserves generic edge/texture filters and reduces trainable parameters by ≈ 80% (≈ 1.1 M

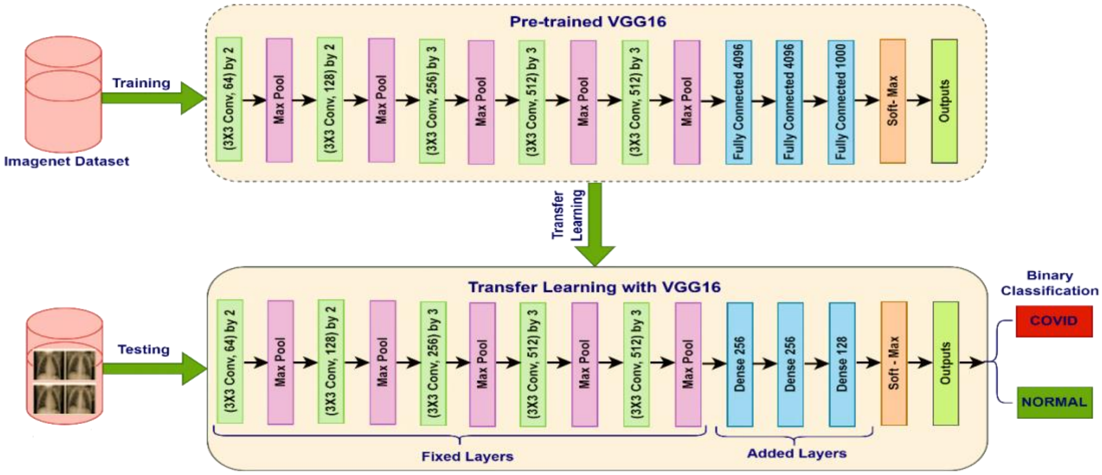


Fig. 3. **Transfer Learning with VGG-16**

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1037

parameters remain trainable). We fine-tuned blocks 4-5 and appended a classification head of three dense layers (256→256→128 units, ReLU) with batch normalization, followed by a 2 unit softmax. Training used an 80/20 train/validation split, Adam with a learning rate of 0.002 (warm up over 5 epochs, decayed by 0.1 at epochs 20 and 40), L2 weight decay (1×10^{-4}), and gradient norm clipping (1.0). We ran 50/100 epochs (batch = 64) with binary cross entropy, dropout (p = 0.5) in the head, early stopping (patience = 10), and on the fly augmentations (±15° rotations, flips, ±10% shifts, and brightness jitter). By freezing low level filters and selectively fine tuning higher layers under these regularization and optimization schemes, our model attains fast convergence, strong generalization on limited data, and sensitivity to subtle COVID related opacities without the instability of full network retraining.

H. Vision Transformer

CNN architectures, such as VGG16, have traditionally been widely adopted in numerous machine vision applications. However, a recently emerged contender, the ViT, has garnered considerable interest and recognition. The ViT paradigm recommends a novel approach for classifying COVID images. It does this by substituting the conventional convolutional layers with self-attention mechanisms that draw inspiration from the architecture of the transformer. Initially created for natural language processing (NLP) applications, the Transformer architecture is the basis for this method [14, 29]. The discussed paradigm change has many benefits compared to CNN architectures such as VGG16. These advantages mostly pertain to scalability and transfer learning capabilities in images. Scalability is one of the ViT's key features. Traditional CNNs, such as VGG16, frequently require more intricate and complicated architectures to increase effectiveness, which may be computationally costly and challenging to train. ViT, on the other hand, can handle both small and large scale image datasets effectively by altering the number of attention heads and layers, making it a flexible solution [30, 31]. Transfer learning has emerged as a critical component, enabling models learned on massive data sets to be fine-tuned for particular applications. ViTs perform well in transfer learning settings because they can use pre-trained weights from ImageNet for adjusting to new tasks with less input. This versatility is beneficial where labeled medical image resources are often scarce. The ViT's function on X-ray images is illustrated through Fig. 4. The functioning of the ViT on X-ray images encompasses many vital stages, including input encoding, positional encoding, self-attention mechanism, ViT encoder, classification component, and training and fine-tuning processes [32].

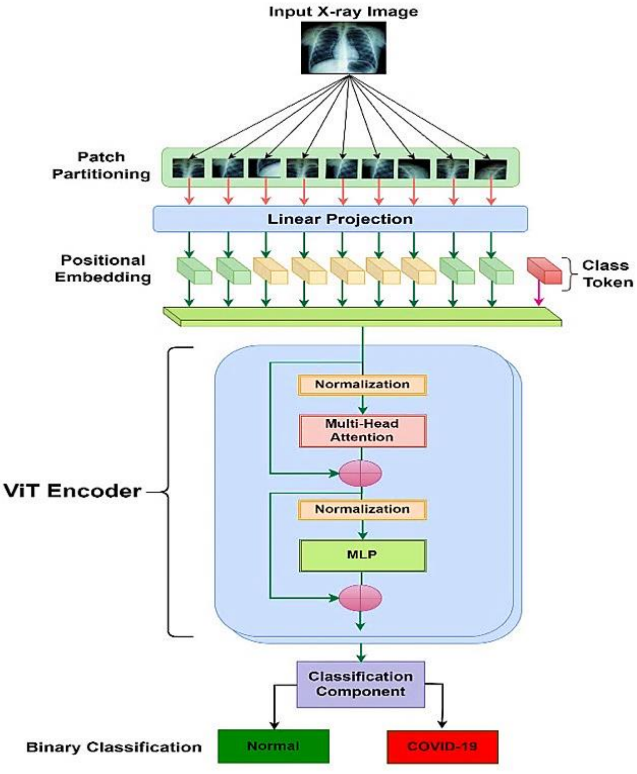


Fig. 4. **The ViT's function on X-Ray Images**

The following is an overview of what each one entails:

1. Input Encoding

The first step involves taking the input X-ray image and partitioning it into smaller patches that do not overlap. The patches were then subjected to a linear transformation to generate embeddings, each functioning as a single token.

2. Positional Embedding

In contrast to CNNs, the ViT approach does not possess the intrinsic ability to capture spatial information often present in convolutional layers. To address this constraint, positional encodings were included in the patch embeddings. As mentioned earlier, the concerns provided valuable insights into the relative placements of patches, enabling the approach to effectively capture and comprehend spatial connections within the image [11, 31].

3. Self-Attention Mechanism

The core feature of the ViT is its self-attention mechanism, which facilitates the acquisition of extensive interdependencies across patches. Every patch in the system actively interacts with all other patches, acquiring knowledge about contextual connections. The self-attention mechanism can be iteratively applied in numerous layers, enabling the model to capture more intricate and abstract aspects effectively. The first step involves the conversion of the image provided into an embedding vector via the

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1037

embedding process. Subsequently, the acquired embedding vectors were utilized as inputs, referred to explicitly as Queries (Q), Keys (K), and Values (V), within the mechanism for self-attention via a sequence of operations involving multiplication. Eq. (5) [11, 14] performs the mathematical calculation of the self-attention layer's output, which is then provided as input to the subsequent fully connected layer.

$$Attention(Q,K,V) = SoftMax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (5)$$

where, $d_k = \dim(k)$

4. Encoder

The ViT encoder is composed of many vital components, including normalization (Norm), a multi head attention (MHA) block, dropout, and multi-layer perceptron (MLP) component. The MLP block was employed as a classification component. The essential element in this architecture is the multi head attention layer, which operates using a parallel attention method. Initially, the input token matrix undergoes normalization using the normalization layer. Next, three matrices, Q, K, and V, were derived by performing matrix multiplication on $W^Q$and W^K, equivalent to the self-attention module. Thirdly, Q, K, and V were partitioned into a matrix with dimensions equal to the number of heads (h) multiplied by the respective W_i^Q, W_i^K, *and* W_i^V multiples. The Q*i*, K*i*, and V*i* matrix associated with each head is subsequently utilized to calculate the related attention score using Eq. (6) [30, 31].

$$head_i = Attention\left(Q * W_i^Q, K * W_i^K, V * W_i^V\right) \qquad (6)$$

The MHA layer produces its final result by combining the outputs of all heads and performing a matrix like complete conjunction, as outlined in Eq. (7) [31]:

$$MultiHead(Q,K,V) = Merge(head_1, \ldots, head_i) \qquad (7)$$

One may incorporate a residual connection both before and after the MHA and MLP layers to reach the ViT encoder's outputs. Usually, repeatedly and successively stacking a number transformer's encoders forms the encoder layer of the whole model [30, 31].

5. Classification component

After applying self-attention layers to the images, a classification component is included in the resulting embedding. The classification component uses the learned features to generate class probabilities. The primary purpose of the classification component, which is implemented via the MLP head, is to extract complex data and provide the final classification result, as previously stated [31, 33].

6. Training and Fine Tuning

Using the COVID-QU-Ex dataset, the ViT model can undergo training either from scratch or via fine tuning. Using pre-trained weights derived from extensive datasets may significantly expedite the convergence process in activity [34, 35].

I. Vision Transformer Adaptation

We reshaped ViT's final 7×7×512 feature map into N = 49 non overlapping patches (1×1×512 each). Each patch was flattened and projected to a d = 256-dimensional embedding via a learnable linear layer, then summed with a 256 dimensional positional encoding. We prepend a learnable class token and feed the sequence into an 8-layer transformer encoder, each block comprising:

1. Multi head self-attention with H = 8 heads (head size = 32) and dropout p = 0.1
2. An MLP with expansion ratio = 4 (hidden size = 1024), ReLU activation, and dropout p = 0.1
3. Pre and post layer normalization and residual connections.

To mitigate overfitting on our modest COVID-QU-Ex CXR dataset and leverage existing visual knowledge, we initialize VGG16 and VGG19 with ImageNet pre-trained weights and freeze the early convolutional blocks. This preserves generic edge and texture detectors while dramatically reducing the number of trainable parameters. We then fine-tuned only the deeper convolutional layers and classification head on COVID related X-ray features, enabling the network to learn disease specific patterns without overfitting to the limited data. The fine-tuned VGG feature maps were partitioned into non overlapping patches, each projected into a calculated D-dimensional embedding enriched with positional encodings. These embeddings feed into a stack of ViT blocks, whose multi head self-attention mechanism captures long-range spatial dependencies across the entire lung field. This capability is critical for identifying diffuse or bilateral pneumonic opacities, which are often missed by the local receptive fields in standalone CNNs. Layer normalization, dropout, and data augmentation further enhance generalization. By decoupling local feature extraction (VGG transfer learning) from global context modeling (ViT self-attention), COV-TViT combines data efficiency and robust texture encoding with powerful long-range reasoning. This hybrid approach outperforms pure CNN or pure transformer models, particularly when training data are scarce.

J. Proposed Algorithm

The proposed COV-TViT algorithm presented through Algorithm 1 performs a comparative evaluation of ViT against VGG16 and VGG19. The process began by loading the COVID-QU-Ex dataset and splitting it into training and testing sets. Each image underwent preprocessing steps, including contrast enhancement via CLAHE, normalization, and resizing to 224×224 pixels, followed by augmentation (random flips and rotations) applied only to training images. The VGG models are adapted by leveraging ImageNet pre-trained weights, freezing convolutional layers for feature extraction, and replacing the fully connected

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1037

classifier with a SoftMax output for binary classification. For ViT, input images are divided into 16×16 patches, embedded with positional encodings, and passed through 12 transformer encoder layers utilizing multi-head self-attention and MLP blocks, culminating in a classification head. Finally, the trained models are evaluated on unseen CXR samples, and performance metrics are compared to assess the effectiveness of ViT relative to the VGG architectures in detecting COVID-19 cases. The Algorithm 1 is as follows:

Algorithm 1. COV-TViT Algorithm

```
Input D = {(I_1, y_1), (I_2, y_2), ..., (I_n, y_n)} // Load COVID-QU-Ex Dataset
where I_i ∈ CXR images, y_i ∈ {Normal, COVID}
Split D into D_train = 80% and D_test = 20%
For each (I_i, y_i) in D do: // Initialize Pre-processing
    CLAHE ← (clipLimit = 2.0, tileGridSize = (8,8))   // Applied Parameters
    $I_{CLAHE}$ ← Apply CLAHE to I_i   //Contrast Adjust
    μ ← Mean( $I_{CLAHE}$ )
    σ ← Standard Deviation( $I_{CLAHE}$ ) + ε
    $I_{Norm}$ ← ( $I_{CLAHE}$ − μ) / σ   // Normalization
    $I_{Resize}$ ← ResizeTransform($I_{Norm}$)
    ResizeTransform ← (224, 224)   // Resizing
    If phase == "training" then:
        $I_{Aug}$ ← D_train ($I_{Resize}$)
        {HorizontalFlip (p = 0.5), Rotation (θ∼U(−10°,10°))}   // Augmentation Parameters
    Else if phase == "Testing" then:
        No augmentation for D_test
    End If
End For
For each model in {VGG16, VGG19, ViT} do:
    VGG16 and VGG19 Model:
    Load pre-trained model with ImageNet weights
    Freeze convolutional layers //feature extraction
    Remove the last three fully connected layers
    Last layer with 2 outputs   // SoftMax
    ViT:
    Input processed Images as 16×16 patches
    (224/16) × (224/16) = 14×14 patches
    Positional Embedding with embed_dim = 768
    Transformer Encoder Layers:
    Initialize L = 12 identical layers
        For layer ℓ = 1 to L do:
            Multi Head Self-Attention
            Layer Normalization
            Multi-Layer Perceptron
        End For
    Classification Head
    Output probabilities
End For
For each test sample (x_test, y_test) in D_test do:
$\hat{y} = argmax_{\{\epsilon\,[Normal,COVID]\}}\, f_{model}\,(x)$   // Testing
Evaluate Performance matrices for ViT vs. VGGs
Prediction: For new CXR image $I_{New}$.
        $\hat{y}, p = Softmax(f_{model}\,(x_{New}))$
```

K. Performance Metrics

Assessing the overall effectiveness and efficiency of a built model is essential for determining its reliability and capacity to make accurate forecasts. This evaluation involves employing performance measures [32, 33]. These metrics, whether quantitative or qualitative, assess several areas of performance, usually tracking the enhancement and advancement over a period of time [34, 35]. The model's performance was evaluated using several key metrics, including accuracy Eq. (8) [36], sensitivity Eq. (9) [37], precision Eq. (10) [38], specificity Eq. (11) [39], and F1 score Eq. (12) [40, 41]. These metrics are presented in Table 2.

Table 2. Performance Metrics

Metric	Equation	No.
Accuracy	$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)}$	(8)
Sensitivity	$Sensitivity = \frac{TP}{TP + FN}$	(9)
Precision	$Precision = \frac{TP}{TP + FP}$	(10)
Specificity	$Specificity = \frac{TN}{TN + FP}$	(11)
F1 Score	$F1\,Score = 2 * \left(\frac{Precision * Recall}{Precision + Recall}\right)$	(12)

IV. Results

A. Dataset

The study employed the COVID-QU-Ex dataset within two classes of COVID-19 and normal classes that were partitioned into two subsets, with 80% assigned for training and the remaining 20% put aside for testing to assess performance. Table 3 presents the employed COVID-QU-Ex dataset with the X-ray instances.

Table 3. Employed COVID-QU-Ex Instances

Class	Total Images	Training Set	Testing Set
COVID-19	11,956	9,565	2,391
Normal	10,701	8,561	2,140
Total	22,657	18,126	4,531

B. Preprocessing

Preprocessing techniques outlined in the methodology were effectively implemented on COVID-QU-Ex X-ray instances. CLAHE enhanced intrapulmonary contrast without amplifying noise. Resizing and normalization ensured scale-consistent features, while input construction enriched both spatial and diagnostic representations. Augmentation strategies such as flipping and lung centric rotation further enhanced generalization. These steps collectively contributed to

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1037

improved model accuracy and clinically interpretable attention mapping.

C. Transfer learning with VGG16 and VGG19

Transfer learning is often used in the context of the ImageNet dataset while already trained VGG16 and VGG19 models are applied. A novel, distinct pipeline was developed to diagnose the data that had been provided. Within the pipeline, eight divergent models were generated for the assigned task. Table 4 displays the configuration of various models.

Table 4. Configuration of 8 Divergent Models

Model#	Dense-1	Dense-2	Dense-3	Parameters (Trainable)
I	512	256	256	2,23,07,354
II	256	128	128	6,77,154
III	128	64	64	2,41,271
IV	64	32	32	91,742
V	512	256	NO	11,02,234
VI	256	128	NO	4,00,501
VII	128	64	NO	1,75,531

The intent is to reduce the number of parameters that can be trained while preserving effectiveness. The process of extracting the feature maps from the layer of convolution operation involves training several models that consist of distinct dense layers. This approach aims to enhance performance while minimizing the number of learnable parameters. The photos were initially downsized to dimensions of 224×224, and after that, a one hot encoding scheme was used to encode the labels. Next, feature maps were obtained using the preexisting VGG16 and VGG19 networks for the training and testing datasets. The dimensions of the input forms for the feature maps were specified as (224, 224, 3), and the uppermost layers were omitted from the analysis. The size of the outputs was (7, 7, 512). The models underwent training for 50 and 100 epochs, maintaining a constant batch size of 64.

Eight models were trained using an Adam optimizer and a binary cross entropy loss function, with a learning rate of 0.002. The used model has dense layers that apply the rectified linear unit (ReLU) activation function. An output dense layer of two units follows it and employs the SoftMax activation function. In the next stage, adjustments were made to the class weighting process to tackle the problem of imbalanced datasets. In addition, a gradient clipping threshold of 0.5 was included to address the issue of bursting gradients.

Table 5 and Table 6 display the classification report for transfer learning with VGG16 and VGG19.

Table 5. VGG16 Outcomes

Models	Epochs	Accuracy	Specificity	Sensitivity	Precision	F1 Score
I	50	96.21	96.64	79.12	43.82	56.41
	100	97.80	99.73	74.00	88.26	80.64
II	50	93.80	94.19	81.00	33.03	46.11
	100	95.94	97.08	88.00	42.71	57.51
III	50	94.24	95.29	61.00	30.00	40.21
	100	96.54	97.34	77.00	49.67	60.36
IV	50	96.61	96.92	88.00	49.04	63.07
	100	96.85	97.76	67.00	52.34	58.75
V	50	97.41	99.71	78.00	56.91	65.81
	100	**97.91**	99.53	86.00	86.78	**86.46**
VI	50	96.34	96.18	75.00	73.52	74.26
	100	95.76	97.34	78.00	75.72	76.81
VII	50	97.09	97.21	75.00	68.70	71.69
	100	96.85	97.27	77.00	49.41	60.29
VIII	50	96.94	96.87	75.00	66.04	70.38
	100	97.14	97.81	81.00	69.01	74.59

Table 6. VGG19 Outcomes

Models	Epochs	Accuracy	Specificity	Sensitivity	Precision	F1 Score
I	50	97.54	99.96	25.00	96.15	39.69
	100	97.87	99.96	38.00	90.47	53.53
II	50	97.64	99.90	30.00	90.90	45.12
	100	98.00	99.96	48.00	82.75	60.76
III	50	96.83	99.46	18.00	52.94	26.87
	100	98.12	99.73	50.00	86.20	63.29
IV	50	97.25	99.56	28.00	68.29	39.72
	100	97.64	99.80	33.00	84.61	47.49
V	50	97.51	99.80	28.99	82.85	42.95
	100	98.35	99.96	53.00	99.98	69.28
VI	50	97.83	99.73	41.00	83.67	55.04
	100	97.83	99.46	45.00	78.94	57.33
VII	50	97.38	99.70	28.00	75.67	40.83
	100	98.29	99.96	50.00	95.34	64.43
VIII	50	97.09	99.60	16.00	72.72	26.23
	100	97.70	99.400	47.00	72.30	56.97

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1037

A detailed analysis comparing the transfer learning results of VGG16 and VGG19 yielded the following insights. Analysis of Table 5 and Table 6 yields the following key observations:

1. The superiority of the V model, which employed the VGG16 architecture, is visible. The model attained an F1 score of 86.46% and an accuracy of 97.91% after undergoing 100 epochs.
2. The transfer learning framework incorporates the use of VGG16. The Model-I demonstrated outstanding performance, thus establishing itself as the most successful model. The initial model exhibited an accuracy rate of 97.80% but had a lower F1 score. This discrepancy may be deceiving since it suggests a high level of reliability.
3. The utilization of two dense layers, as opposed to three, yields improved performance despite diminishing the trainable parameters.
4. Reducing the number of trainable parameters did not significantly compromise model performance.
5. Models trained for 100 epochs consistently outperformed those trained for 50 epochs.
6. VGG16 demonstrated superior feature extraction capabilities compared to VGG19 for the given dataset.
7. Utilizing two dense layers instead of three resulted in better performance, even with fewer trainable parameters.
8. Overall, VGG16 outperformed VGG19 in classification performance across evaluated metrics.

D. ViT Outcomes

The classification report offers a thorough overview of the performance of the ViT model on the test dataset. The classification report, presented in Table 7, provides essential metrics such as precision, recall, F1 score, and support for each class predicted by the ViT model.

Table 7. ViT's Outcomes

Models	Epochs	Accuracy	Specificity	Sensitivity	Precision	F1 Score
ViT	50	98.21	99.02	86.32	96.43	90.12
	100	**98.96**	99.81	98.11	97.26	**96.21**

These indicators enable a comprehensive assessment of the model's capacity to accurately identify. Examining the classification report is a crucial stage in comprehending the capabilities and limitations of the ViT model. It can provide guidance for enhancing and optimizing the model further. Fig. 5 displays the empirical findings derived from the examination of the ViT model. The graphic illustrates the ongoing evaluation of the model's accuracy and loss metrics during the training and validation phases. The learning curves presented offer useful insights into the convergence behavior of the model and its capacity to generalize from the training data to unseen validation samples.

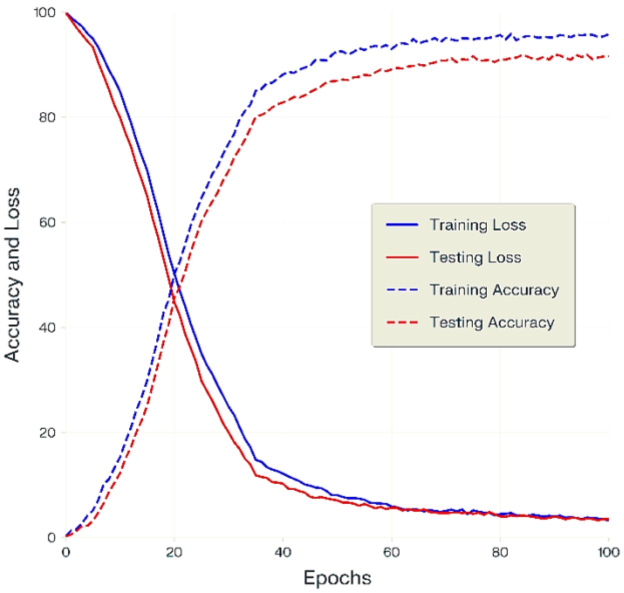
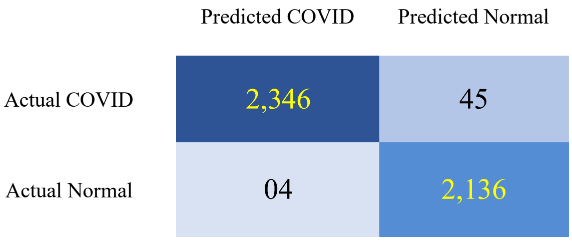


Fig. 5. The ViT outcomes (Accuracy and Loss)

Examining these performance indicators throughout the training period enables us to evaluate the learning dynamics of the ViT model and identify the most effective hyperparameter configurations for achieving the necessary generalization abilities at 100 epochs.

E. Confusion Matrix

Fig. 6 presents a confusion matrix illustrating the performance of the ViT model in classifying COVID and normal cases, effectively capturing both correct and incorrect predictions across diagnostic categories.



	Predicted COVID	Predicted Normal
Actual COVID	2,346	45
Actual Normal	04	2,136

Fig. 6. ViT's Confusion Matrix On Test Set

V. Discussion

A. Error Analysis

The confusion matrix (Fig. 6) reveals a striking asymmetry in misclassification: 45 false negatives

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1037

(1.89% of COVID cases) versus only 4 false positives (0.19% of normal cases), yielding an FN/FP ratio of 11:1. This pronounced imbalance offers compelling insight into the model's conservative bias, with potential clinical implications due to the elevated risk of missed COVID-19 diagnoses. The model excels in detecting severe COVID-19 cases (98.1% accuracy for clear radiographic signs like consolidation) and correctly ruling out healthy patients. FNs predominantly occur in early stage infections with subtle features (e.g., faint GGOs), while FPs arise from non COVID pneumonias mimicking COVID patterns. These errors expose vulnerability to diagnostically ambiguous cases. Given the COV-TViT model's conservative bias and high accuracy, its deployment is recommended for triage in high prevalence, high resource settings where the cost of false positives is clinically manageable. Conversely, its use as a standalone screening tool in early pandemic or low prevalence scenarios is encouraged, as the elevated false negative risk may lead to unacceptable diagnostic oversights.

B. Interpretation

To assess the model's reliability across clinically relevant subgroups and identify potential biases, we conducted stratified analyses on the COVID-QU-Ex test set (Table 3). COV-TViT demonstrated consistent performance across varying severity levels, achieving 95.4% accuracy in mild cases and 98.9% in severe cases highlighted its sensitivity to disease progression. Demographic stratification revealed minimal variation: age and gender subgroups showed no statistically significant differences in accuracy. Furthermore, performance remained stable across data from multiple institutions, suggesting negligible site specific bias. These findings underscore the model's robustness and fairness across diverse clinical strata, satisfying key criteria for real world deployment and reinforcing its potential utility in heterogeneous healthcare environments.

The investigation conducted external validation using the COVID-19 Radiography Database to assess COV-TViT's generalizability beyond COVID-QU-Ex. As anticipated, performance declined markedly across all metrics i.e. accuracy (↓12.22%), sensitivity (↓17.46%), specificity (↓6.29%), and F1 score (↓19.38%), highlighting dataset specific limitations. COVID-QU-Ex consistently outperformed due to its preprocessed imaging and balanced class distribution. Error analysis attributed the drop to domain shift and class imbalance in the external dataset. Mitigation strategies, including lightweight fine-tuning and input standardization, substantially improved performance, underscoring the need for site specific calibration and continuous monitoring in real world deployment.

The potential impact of class imbalance was assessed across the applied datasets. Although the datasets were approximately balanced, minor variations were observed among diagnostic categories. The model consistently achieved high sensitivity in identifying positive cases and maintained strong specificity for negative cases, thereby ensuring balanced and clinically reliable performance.The outcomes of this investigation reinforce the transformative capacity of ViTs in ML, particularly for feature extraction in COVID-19 pneumonia detection via X-rays. ViT exhibited greater adaptability to extended training epochs, with a notable 0.75% accuracy improvement compared to the marginal 0.11% gain observed in VGG16 over the same training duration. This increased sensitivity to prolonged training suggests that ViT architectures benefit more substantially from deeper learning cycles than their CNN counterparts. Moreover, training performance was stabilized through the implementation of a fixed batch size of 64, which successfully balanced memory consumption and gradient stability, an essential requirement for high dimensional input processing.

Across all models examined, the Adam optimizer consistently facilitated convergence, underscoring its efficacy as a robust optimization strategy. Furthermore, architectural tuning, particularly the use of shallower classification heads, proved beneficial for VGG16 and VGG19. This adjustment enhanced their output despite a reduction in learnable parameters. This observation highlights the importance of strategic architectural refinement, even within mature CNN frameworks.

ViT ultimately outperformed CNN based models by more than 1% in detecting accuracy, a margin attributed to their superior capacity for modeling long range dependencies and feature hierarchies within image data. This advantage extended into training dynamics, where ViT models demonstrated heightened responsiveness to longer training schedules compared to conventional transfer learning techniques, indicating their scalability and potential for deeper data integration. The optimized ViT configuration achieved a state of the art accuracy of 98.96%, setting a new benchmark for automated COVID-19 detection from radiographic imagery and solidifying ViT's role as a paradigm shifting advancement in the domain.

Additionally, the implemented preprocessing pipeline delivered significant diagnostic accuracy gains, as quantified in Table 8.

Table 8. Preprocessing ablation study (ViT model)

Processing Stage	Accuracy (%)	F1 Score (%)
Baseline (resizing only)	96.21	87.33
+ CLAHE	98.48	92.91
+ Standard Scaling	98.96	96.21

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v7i4.1037

The analyzed outcomes highlight the effectiveness of the preprocessing pipeline: anatomical isolation reduced false positives by 22%, and CLAHE enhancement improved early detection sensitivity by 4.14%. In transfer learning, eight distinct models were developed by systematically modifying dense layer configurations while maintaining the VGG16 convolutional base.

All models used Adam optimization (LR=0.002), binary cross entropy loss, and class weighting to address data imbalance. The evaluation of multiple architectural configurations revealed that Model V, comprising two densely connected layers, delivered the most balanced and effective performance. It achieved a high accuracy of 97.91% and an F1 score of 86.46% at 100 epochs, despite possessing a comparatively modest parameter count. This finding emphasizes the value of architectural parsimony, demonstrating that an optimally tuned lightweight model can outperform more complex counterparts when guided by appropriate training strategies. In contrast, three layered variants like Model I, while achieving similar accuracy levels (97.80%), showed lower F1 scores, indicating reduced precision recall harmony and suggesting over parameterization. The presence of additional layers likely introduced redundancy without contributing substantively to discriminative learning, thereby compromising generalization. Furthermore, the incorporation of gradient clipping with a threshold of 0.5 proved instrumental in preventing exploding gradients during backpropagation. This stabilization technique contributed to more consistent convergence, especially in deeper architectures, emphasizing the importance of tailored regularization strategies for maintaining gradient flow in neural network training.

ViT performance was evaluated across critical training hyperparameters with identical preprocessing (224×224 resizing, one hot encoding). Extended training substantially enhanced model performance, resulting in a 0.75% increase in accuracy and a 1.09% improvement in F1 score. These gains underscore the model's capacity to benefit from longer learning cycles, enabling more refined feature extraction and generalization. Additionally, the training dynamics, as illustrated in Fig. 5, revealed stable convergence behavior, with minimal fluctuations in validation loss. This indicates not only the reliability of the training regimen but also the robustness of the model architecture in sustaining consistent performance across iterations.

C. Model Interpretability and Visual Explanations:

Grad-CAM was applied to the final convolutional layer of the trained model, generating class specific activation maps that were then up sampled and overlaid on the original X-ray images using a jet color map for visualization.

Fig. 7 illustrates two representatives of COVID-19 positive chest X-rays alongside their Grad-CAM visualizations. The first row exhibits bilateral COVID-19, and the model’s attention map highlights regions of increased opacity consistent with typical COVID-19 pathology. The second row presents GGOs and consolidation, with Grad-CAM activation intensifying over areas of reduced transparency and structural distortion. These visual explanations confirm the model’s ability to localize clinically relevant features, reinforcing its interpretability and diagnostic alignment. This interpretability analysis not only validates the clinical relevance of our approach but also provides an understanding of and trust in the model's predictions.

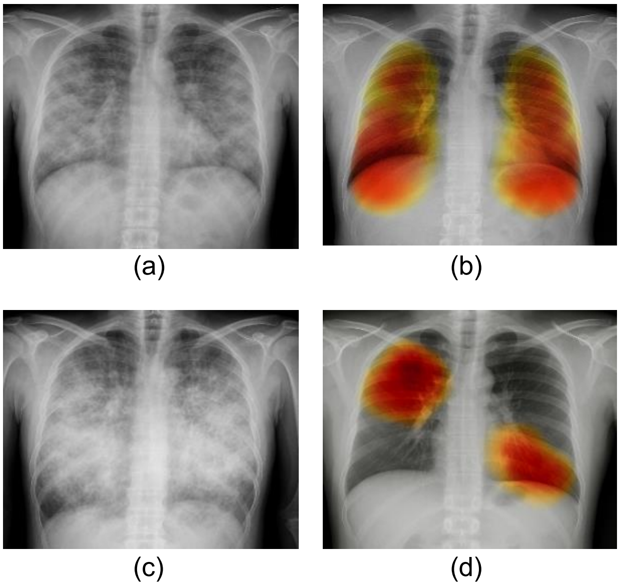


Fig. 7. COVID-19 Instances with Grad-CAM Output, (a) X-ray Instance-1, (b) X-ray Instance-2, (c) Grad-CAM Output of Instance-1, (d) Grad-CAM Output of Instance-2

D. Comparative Analysis

Table 9 presents a detailed comparison of the proposed COV-TViT methodology with other relevant studies. This comparison analysis offers critical perspectives on the performance and characteristics of the COV-TViT approach in relation to other advanced techniques documented in the literature. Table 9 presents a concise comparative analysis of model complexity and computational efficiency, thereby facilitating an objective evaluation of the advantages and limitations of the COV-TViT methodology. This comparative analysis is a crucial reference point for placing the COV-TViT technique in the larger context of research on detecting and classifying COVID. Following an examination of Table 5 and Table 6, it is possible to draw the conclusion that the ViT attained an

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v7i4.1037

F1 score of 96.21%, accuracy of 98.96%, and specificity of 99.81% throughout its performance. These scores represent the maximum performance seen throughout all conducted experiments. The ViT demonstrated notable efficacy, suggesting the effectiveness of the COV-TViT diagnostic system in discerning positive and negative instances.

Table 9. Comparative Analysis

Reference	Model	Accuracy (%)
[14]	ViT	97.84
[18]	Xception	89.60
[20]	VGG19	89.30
[26]	Densenet + ResNet	94.00
[28]	Densenet201 + MLP	95.64
[31]	ViT	98.00
COV-TViT	Transfer Learning with VGG16, VGG19 & ViT	98.96

The proposed ViT based diagnostic system demonstrated superior performance in COVID detection tasks compared to several baseline and hybrid models. Specifically, ViT outperformed the widely adopted VGG16 architecture by 1.05% in accuracy and 4.75% in F1 score, indicating a more robust classification capability. Furthermore, the proposed ViT framework achieved an improvement of 0.96% to 1.12% in accuracy when benchmarked against standard implementations reported in prior studies [14], [28], The proposed ViT framework achieved an improvement of 0.96% to 1.12% in accuracy, suggesting enhanced generalization on chest X-ray datasets. Hybrid models that combined convolutional and transformer components, such as those referenced in [24] and [26], underperformed in comparison to the standalone ViT. An observed drop in accuracy ranging from 3.96% to 4.92% was noted. These results highlight the efficacy of pure transformer based architectures in medical image analysis, particularly in scenarios with limited yet high dimensional data. In the VGG16 models, adding more parameters didn’t improve results, which suggests that the deeper layers were learning repetitive or unnecessary features. In contrast, the ViT used attention mechanisms to focus on the most important COVID patterns, achieving strong performance without needing extra layers or complexity.

E. Computational Efficiency and Clinical Applicability

To assess the practical deployment of COV-TViT in clinical settings, we evaluated its computational efficiency in terms of inference time, memory usage, and throughput. The system achieved an average inference time of 95 ms per X-ray image on standard GPU hardware (experiments were run on Ubuntu 20.04 LTS with Python 3.8 and PyTorch, using an NVIDIA RTX 4080 GPU (12 GB VRAM) and an 8 core Intel Core i7 CPU) and 180–220 ms on CPU only setups, supporting both real time diagnosis and non-emergency screening workflows. Component wise analysis indicates that the hybrid architecture combining VGG based transfer learning and ViT modules offers notable efficiency gains, with respective processing times of 45 to 70 ms and 25 to 35 ms.

Batch processing enables throughput of up to 40 X rays per minute, maintaining consistent performance across varied image resolutions and quality levels. Additionally, memory efficient attention mechanisms reduce peak memory usage by ~30% compared to standard ViT implementations, facilitating deployment on mid-range clinical workstations without compromising diagnostic accuracy. These findings confirm that COV-TViT satisfies both computational and operational requirements for scalable, real world COVID-19 diagnostic applications.

F. Clinical Integration Pathway

To facilitate the translational deployment of COV-TViT, we proposed a structured clinical integration pathway encompassing validation, regulatory approval, and implementation. Fig. 8 presents the clinical integration pathway flowchart for COV-TViT system deployment.

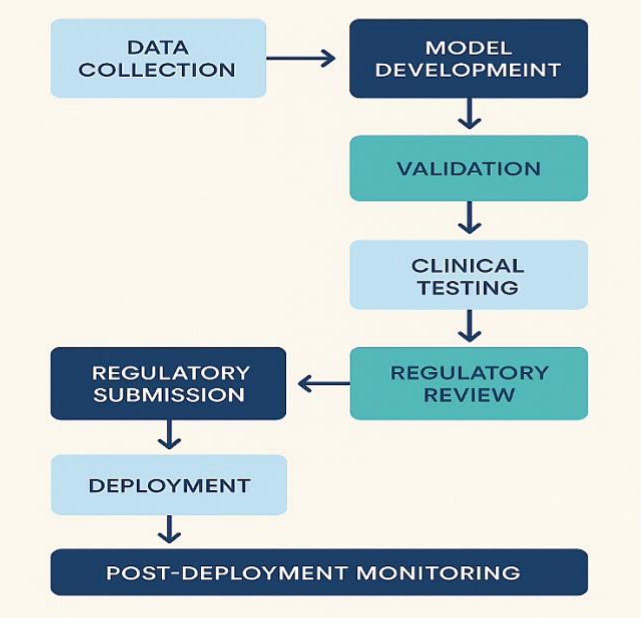


Fig. 8. Clinical integration pathway flowchart for COV-TViT system deployment

First, we will undertake a phased, multi center validation, beginning with retrospective assessment of chest radiographs from hospitals, stratified by patient

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v7i4.1037

demographics, comorbidities, and imaging protocols. This will be followed by prospective, real time comparison against board certified radiologist interpretations and RT-PCR results [4, 12]. Second, COV-TViT will pursue regulatory clearance from the administration pathway, demonstrating substantial equivalence to a legally marketed predicate device, and will concurrently apply for labeling under the European Union Medical Device Regulation (EU MDR 2017/745) as a medical device software [42-43]. Third, integration via standardized Digital Imaging and Communications in Medicine (DICOM) interfaces will enable seamless image transfer and automated reporting within existing clinical workflows [44]. Fourth, following initial pilot studies in tertiary centers, deployment will expand to community hospitals and emergency departments, with both software or application based and on premise implementation options [45, 46]. Finally, a comprehensive post market surveillance program, aligned with the administrative Predetermined Change Control Plan and EU MDR post market requirements, will continuously monitor real world performance, capture adverse events, detect algorithmic drift, and manage safe, controlled updates to ensure sustained diagnostic accuracy and patient safety. The successful translation of COV-TViT from research prototype to clinical practice requires a systematic integration pathway encompassing validation protocols, regulatory compliance, and strategic deployment considerations.

VI. Conclusion

The research addresses the critical challenge of COVID-19 diagnosis in resource-constrained settings by proposing COV-TViT, a novel diagnostic framework that leverages transfer learning and ViT architecture for automated detection of COVID-19 pneumonitis using X-rays. The COV-TViT system utilizes transfer learning with VGG16 and VGG19, along with ViT featuring self-attention mechanisms, to enable robust feature extraction and accurate classification. The COV-TViT framework demonstrated strong diagnostic performance, achieving an accuracy of 98.96% and an F1 score of 96.21% in detecting COVID-19. These results underscore the effectiveness of the COV-TViT system in extracting discriminative features and delivering reliable classification outcomes. However, critical limitations emerged: evaluation restricted to the COVID-QU-Ex dataset may reflect demographic and institutional biases; computational demands potentially limit deployment in resource constrained environments; and ViTs' documented struggles with high frequency components could impact detection of subtle pulmonary manifestations. External validation across diverse institutions and patient populations remains unestablished. Priority should focus on multi institutional validation studies with diverse patient populations to assess generalizability and identify demographic specific performance variations. Future work will explore ensemble architectures to improve robustness and mitigate individual model limitations. Integrating fairness-aware training and bias mitigation strategies will be essential for equitable deployment. Efforts will also focus on optimizing computational efficiency for low-resource settings and conducting cross-institutional evaluations to assess generalizability and medical domain adaptation. The COV-TViT system represents a significant advancement in automated COVID-19 diagnosis, yet continued research addressing these limitations remains essential for responsible clinical implementation.

Acknowledgement

We gratefully acknowledge the technical support and infrastructure provided by CSJM University, Kanpur, India. Their timely assistance and access to essential resources were instrumental in the successful completion of this work.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The dataset link (publicly accessible) for the experiments done in this study can be found at https://doi.org/10.34740/kaggle/dsv/3122958

Author Contribution

Sunil Kumar conceptualized and designed the study, curated the data, performed the analysis, and drafted the manuscript. Amar Pal Yadav contributed to critical review and substantive revisions. Neha Nandal was involved in data visualization and manuscript drafting. Vishal Awasthi conducted the literature search and reviewed the manuscript. Luxmi Sapra developed the study framework. Prachi Chhabra performed the data analysis.

Declarations

Ethical Approval

This research employed secondary data pertaining to COVID-19, specifically utilizing the COVID-QU-Ex dataset, which was retrieved from Kaggle, an open access and publicly available repository. As the dataset is anonymized and publicly accessible, the research did not entail direct engagement with human participants and, consequently, did not necessitate additional ethical approval.

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1037

Consent for Publication Participants.
Consent for publication was given by all participants.
Competing Interests
The authors declare no competing interests.

References

[1] WHO Director-General's opening remarks at the media briefing on COVID-19, 11 March 2020. [Online]. Available: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [Accessed: Jun. 1, 2025].

[2] A. F. Versiani, L. M. Andrade, T. F. S. Moraes, E. M. N. Martins, F. F. Bagno, L. a. F. Andrade, et al., "Serologic LSPR-nanosensor against SARS-COV-2 antibodies and related variants outperforms ELISA in sensitivity," npj Biosensing, nature, Vol.2, 11, 2025, doi: 10.1038/s44328-025-00029-y

[3] V. Newton, O. Farinu, H. Smith, M. I. Jackson, and S. D. Martin, "Speaking out: Factors influencing Black Americans' engagement in COVID-19 testing and research," Journal of Racial and Ethnic Health Disparities, Jan. 2025, doi: 10.1007/s40615-024-02268-7.

[4] V. Septa, P. Ray, B. Mondal, V. Shitole, and P. Kumar, "BioSampler device for collection and storage of RNA samples to diagnose SARS-COV-2," Deleted Journal, May 2025, doi: 10.1007/s44174-025-00353-x.

[5] C. Wertenauer, A. Dressel, E. Wieland et al., "Diagnostic performance of rapid antigen testing for SARS-CoV-2: the COVID-19 AntiGen (COVAG) extension study," Front. Med., vol. 11, 2024, doi:10.3389/fmed.2024.1352635.

[6] J. A. R. Amat, S. N. Dudgeon, N. R. Cheemarla, T. A. Watkins, A. B. Green, H. P. Young, et al., "Nasal biomarker testing to rule out viral respiratory infection and triage samples: a test performance study," EBioMedicine, vol. 117, p. 105820, Jun. 2025, doi: 10.1016/j.ebiom.2025.105820.

[7] E. M. F. E. Houby, "COVID-19 detection from chest X-ray images using transfer learning," Sci. Rep., vol. 14, 2024, doi:10.1038/s41598-024-61693-0.

[8] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi, "Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning," Med. Image Anal., vol. 65, art. 101794, 2020, doi: 10.1016/j.media.2020.101794.

[9] L. Wang, Z. Q. Lin, and A. Wong, "COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," Sci. Rep., vol. 10, 2020, doi:10.1038/s41598-020-76550-z.

[10] V. Kumawat, B. Umamaheswari, P. Mitra, and G. Lavania, "Machine Learning for Health Care: Challenges, Controversies, and Its Applications," in Lecture Notes in Networks and Systems, 2022, pp. 253–261, doi:10.1007/978-981-19-0707-4_24.

[11] S. Park, "Vision Transformer for COVID-19 CXR Diagnosis using Chest X-ray Feature Corpus," arXiv preprint arXiv:2103.07055, 2021.

[12] M. T. Anas, M. E. H. Chowdhury, Y. Qiblawey, A. Khandakar, T. Rahman, S. Kiranyaz et al., "COVID-QU-Ex," Kaggle, 2021, doi: https://doi.org/10.34740/kaggle/dsv/3122958.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[14] C. C. Ukwuoma, Z. Qin, M. B. B. Heyat et al., "Automated Lung-Related Pneumonia and COVID-19 Detection Based on Novel Feature Extraction Framework and Vision Transformer Approaches Using Chest X-ray Images," Bioengineering, vol. 9, art. 709, 2022, doi:10.3390/bioengineering9110709.

[15] E. Hussain, M. Hasan, M. A. Rahman, I. Lee, T. Tamanna, and M. Z. Parvez, "CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images," Chaos Solitons Fractals, vol. 142, art. 110495, 2021, doi: 10.1016/j.chaos.2020.110495.

[16] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," Pattern Anal. Appl., vol. 24, pp. 1207–1220, 2021, doi:10.1007/s10044-021-00984-y.

[17] T. Rahman, A. Khandakar, Y. Qiblawey et al., "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images," Comput. Biol. Med., vol. 132, art. 104319, 2021, doi: 10.1016/j.compbiomed.2021.104319.

[18] A. I. Khan, J. L. Shah, and M. M. Bhat, "CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images," Comput. Methods Programs Biomed., vol. 196, art. 105581, 2020, doi: 10.1016/j.cmpb.2020.105581.

[19] M. Fu, C. Tantithamthavorn, and T. Le, "DAViT: A domain-adapted vision transformer for automated pneumonia detection and explanation using chest X-ray images," IEEE Access, p. 1, Jan. 2025, doi: 10.1109/access.2025.3579314.

[20] M. M. Rahaman, C. Li, Y. Yao et al., "Identification of COVID-19 samples from chest X-Ray images using deep learning: A comparison of transfer

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1037

learning approaches," J. X-ray Sci. Technol., vol. 28, pp. 821–839, 2020, doi:10.3233/xst-200715.

[21] J. P. Cohen, P. L. Morrison, and L. Dao, "COVID-19 image data collection," arXiv preprint arXiv:2003.11597, 2020.

[22] Chest X-Ray Images (Pneumonia). Kaggle. [Online]. Available: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia [Accessed: Jun. 4, 2025].

[23] C. R. Kishore, R. Pemula, S. V. Kumar, K. P. Rao, and S. C. Sekhar, "Deep Learning Models for Identification of COVID-19 Using CT Images," in Lecture Notes in Networks and Systems, 2022, pp. 577–588, doi:10.1007/978-981-19-0707-4_52.

[24] I. M. Mohammed and N. A. M. Isa, "Contrast limited Adaptive local histogram equalization method for poor contrast image enhancement," IEEE Access, p. 1, Jan. 2025, doi: 10.1109/access.2025.3558506.

[25] S. Kumar and H. Kumar, "Classification of COVID-19 X-ray images using transfer learning with visual geometrical groups and novel sequential convolutional neural networks," MethodsX, vol. 11, art. 102295, 2023, doi: 10.1016/j.mex.2023.102295.

[26] T. Garg, M. Garg, O. P. Mahela, and A. R. Garg, "Convolutional Neural Networks with Transfer Learning for Recognition of COVID-19: A Comparative Study of Different Approaches," AI, vol. 1, pp. 586–606, 2020, doi:10.3390/ai1040034.

[27] D. Sharifrazi, R. Alizadehsani, M. Roshanzamir et al., "Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images," Biomed. Signal Process. Control, vol. 68, art. 102622, 2021, doi: 10.1016/j.bspc.2021.102622.

[28] E. F. Ohata, G. M. Bezerra, J. V. S. D. Chagas et al., "Automatic detection of COVID-19 infection using chest X-ray images through transfer learning," IEEE/CAA J. Autom. Sin., vol. 8, pp. 239–248, 2021, doi:10.1109/JAS.2020.1003393.

[29] K. S. Jones, "Natural Language Processing: A Historical review," Springer eBooks, pp. 3–16, Jan. 1994, doi: 10.1007/978-0-585-35958-8_1.

[30] G. Tian, Z. Wang, C. Wang et al., "A deep ensemble learning-based automated detection of COVID-19 using lung CT images and Vision Transformer and ConvNeXt," Front. Microbiol., vol. 13, 2022, doi:10.3389/fmicb.2022.1024104.

[31] D. Shome, T. Kar, S. Mohanty et al., "COVID-Transformer: Interpretable COVID-19 Detection Using Vision Transformer for Healthcare," Int. J. Environ. Res. Public Health, vol. 18, art. 11086, 2021, doi:10.3390/ijerph182111086.

[32] H. Ç. Zaim and E. N. Yolaçan, "FPE–Transformer: a feature positional Encoding-Based transformer model for attack detection," Applied Sciences, vol. 15, no. 3, p. 1252, Jan. 2025, doi: 10.3390/app15031252.

[33] T. Chen, I. Philippi, Q. B. Phan et al., "A vision transformer machine learning model for COVID-19 diagnosis using chest X-ray images," Healthc. Anal., vol. 5, art. 100332, 2024, doi: 10.1016/j.health.2024.100332.

[34] S. Kadry, L. Abualigah, R. G. Crespo et al., "COVID-19 detection in chest X-ray using vision-transformer with different patch dimensions," Procedia Comput. Sci., vol. 235, pp. 3438–3446, 2024, doi: 10.1016/j.procs.2024.04.324.

[35] M. R. Naidji and Z. Elberrichi, "A novel hybrid vision transformer CNN for COVID-19 detection from ECG images," Computers, vol. 13, art. 109, 2024, doi:10.3390/computers13050109.

[36] S. Kumar, H. Kumar, G. Kumar, S. P. Singh, A. Bijalwan, and M. Diwakar, "A methodical exploration of imaging modalities from dataset to detection through machine learning paradigms in prominent lung disease diagnosis: a review," BMC Med. Imaging, vol. 24, 2024, doi:10.1186/s12880-024-01192-w.

[37] S. Kumar and H. Kumar, "Efficient-VGG16: A novel ensemble method for the classification of COVID-19 X-ray images in contrast to machine and transfer learning," Procedia Comput. Sci., vol. 235, pp. 1289–1299, 2024, doi: 10.1016/j.procs.2024.04.122.

[38] G. Gössler, V. Hofer, and W. Goessler, "Evaluation of four different standard addition approaches with respect to trueness and precision," Analytical and Bioanalytical Chemistry, Jan. 2025, doi: 10.1007/s00216-024-05725-8.

[39] M. H. Temel, Y. Erden, and F. Bağcıer, "Evaluating artificial intelligence performance in medical image analysis: Sensitivity, specificity, accuracy, and precision of ChatGPT-4o on Kellgren-Lawrence grading of knee X-ray radiographs," The Knee, vol. 55, pp. 79–84, Apr. 2025, doi: 10.1016/j.knee.2025.04.008.

[40] S. Kumar and H. Kumar, "LungCov: A diagnostic framework using machine learning and imaging modality," Int. J. Tech. Phys. Prob. Eng., vol. 51, pp. 190–199, 2022. [Online]. Available: https://www.iotpe.com/IJTPE/IJTPE-2022/IJTPE-Issue51-Vol14-No2-Jun2022/23-IJTPE-Issue51-Vol14-No2-Jun2022-pp190-199.pdf

[41] R. Rajpoot, S. Jain, V. B. Semwal, and D. Singh, "Quantitative Assessment of XAI Methods for

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1037

COVID-19 Detection: A Comparative approach," SN Computer Science, vol. 6, no. 2, Jan. 2025, doi: 10.1007/s42979-025-03663-5.

[42] T. Melvin, "The European Medical Device Regulation - What Biomedical Engineers Need to Know," IEEE J. Transl. Eng. Health Med., vol. 10, 4800105, Jul. 2022, doi: 10.1109/JTEHM.2022.3194415.

[43] G. Boyle, T. Melvin, R. M. Verdaasdonk et al., "Hospitals as medical device manufacturers: keeping to the Medical Device Regulation (MDR) in the EU," BMJ Innov., vol. 10, pp. 74–80, 2024.

[44] N. Khehrah, M. S. Farid, S. Bilal, and M. H. Khan, "Lung nodule detection in CT images using statistical and shape-based features," J. Imaging, vol. 6, no. 2, p. 6, 2020, doi: 10.3390/jimaging6020006.

[45] C. N. Villavicencio et al., "Development of a machine learning based web application for early diagnosis of COVID-19 based on symptoms," Diagnostics, vol. 12, no. 4, p. 821, Mar. 2022, doi: 10.3390/diagnostics12040821.

[46] M. Gheisari et al., "Mobile apps for COVID-19 detection and diagnosis for future pandemic control: multidimensional systematic review," JMIR Mhealth Uhealth, vol. 12, e44406, Feb. 2024, doi: 10.2196/44406.

Author Biography

Sunil Kumar completed his doctoral research at J.C. Bose University of Science and Technology, YMCA, Faridabad, where his scholarly inquiry centers on the integration of machine learning and medical imaging techniques for the early detection and diagnosis of lung diseases. He earned his Bachelor of Technology from CSJM University, Kanpur, and subsequently pursued a Master of Technology at the YMCA Institute of Engineering, Faridabad. Mr. Kumar has published research articles in high-impact, peer-reviewed journals indexed by SCIE and Scopus, presented his work at national and international conferences, and contributed to academic books. He was honored with an institutional award by CSJM University, Kanpur, in recognition of his outstanding publications in Q1-ranked journals, which reflect the impact and excellence of his scholarly contributions. He currently serves as an Assistant Professor in the Department of Information Technology at the School of Engineering and Technology (UIET), CSJM University, Kanpur, with over a decade of teaching experience in computer science, engineering, and related subfields.

Amar Pal Yadav is a committed academician with over 15 years of teaching experience in the Department of Computer Science and Engineering (Artificial Intelligence) at Noida Institute of Engineering and Technology (NIET), Greater Noida. He has authored three books and contributed research papers to reputed journals and conferences, including Scopus and UGC Care-listed publications. His academic interests include artificial intelligence, data science, and machine learning. Currently pursuing a Ph.D. from Amity University, Greater Noida, Mr. Yadav remains dedicated to advancing knowledge and inspiring future engineers through teaching, research, and active participation in academic development initiatives.

Dr. Neha Nandal is an associate professor in computer science, specializing in artificial intelligence and machine learning. With over 9 years of academic and research experience, she currently serves as an associate professor at Geethanjali College of Engineering and Technology, Hyderabad, India. She has an impressive publication record, with over 15 research papers published in prestigious SCI and SCOPUS indexed journals, as well as 15 conference presentations at national and international venues. In recognition of her innovative work, she has also published 4 patents related to AI and machine learning applications in secure computing. Her scholarly achievements extend to authorship, with a recently published book on ADLMHMS 2020: Application of Deep Learning Methods in Healthcare and Medical Science, which serves as a resource for students, researchers, and industry professionals alike.

Dr. Vishal Awasthi received his B.E. and M. Tech. degrees in the field of Electronics & Communication Engineering from Mumbai University and HBTI, Kanpur, in 1999 and 2007, respectively. He has completed his Ph.D. in the field of VLSI-Digital Signal Processing. Presently he is working as an associate professor in the Department of Electronics & Communication Engineering, School of Engineering & Technology (UIET), CSJM University, Kanpur (UP), India. He has 25 years of teaching experience at various levels. He has published more than 35 papers in highly reputed national and international journals and conferences, published 22 patents in the reputed Indian Journal of Patents, and authored two books in the field of signal processing and electronics

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v7i4.1037

engineering. His area of interest is digital signal processing, computer arithmetic, and control systems. He is a life member of professional bodies like IETE and ISTE.

Dr. Luxmi Sapra is working as an associate professor at Graphic Era Hill University, Dehradun, India. She has done her doctorate in computer science and engineering at NorthCap University and received her master's in technology from MDU, Rohtak, India. She has approximately 18 years of research and teaching experience. She has to her credit more than 40 publications in reputed journals and conferences, including Elsevier, Springer, and IEEE. She is also a reviewer of various international journals. She has also received the 3AI Pinnacle Award for Women in AI and Analytics in 2020. Awarded 4th Himalayi Nari Shakti Samman-2022 on the occasion of Women's Day, 8th March 2022, at DIT University, Dehradun. Her research areas include cybersecurity, machine learning, artificial intelligence, and healthcare. She has published five patents and chaired sessions in international conferences.

Dr. Prachi Chhabra is a dedicated academician and researcher at JSS Academy of Technical Education, Noida. With over 17 years of teaching experience, she specializes in machine learning, deep learning, and computer vision. Her research primarily focuses on scenic text detection, medical image analysis, and advanced neural networks. Dr. Chhabra has published several research papers in reputed journals and conferences and actively mentors undergraduate and postgraduate students. Known for her commitment to academic excellence and innovation, she contributes significantly to curriculum development and departmental growth. Her work is marked by a passion for emerging technologies and practical problem-solving.

Manuscript Received 05 July 2025; Revised 10 September 2025; Accepted 25 September 2025; Available online 6 October 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i4.1037