High-accuracy fine-tuned vision transformer model for diagnosing COVID-19 from chest X-ray images

Tianyi Chen, Ian Philippi, Quoc Bao Phan, Linh Nguyen, Carlo daCunha, Tuy Tan Nguyen*

School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ 86011, USA

Abstract

This research investigates the application of machine learning for diagnosing COVID-19 from chest X-rays. We analyze various popular architectures, including efficient neural networks (EfficientNet), multiscale vision transformers (MViT), efficient vision transformers (EfficientViT), and vision transformers (ViT), on a dataset categorized into COVID, lung opacity, normal, and viral pneumonia. While multiscale models demonstrate a tendency to overfit, our proposed fine-tuning ViT model achieves significant accuracy, reaching 95.79% in four-class classification, 99.57% in a clinically relevant three-class grouping, and similarly high performance in binary classification. Validation through quantitative metrics and visualization solidifies the model's effectiveness. Comparative analysis showcases the superiority of our approach. Overall, these findings showcase the potential of ViT for accurate COVID-19 diagnosis, contributing to the advancement of medical image analysis.

Keywords: COVID-19, chest X-ray, deep learning, vision transformers (ViT), medical applications

1. Introduction

The current reliance on X-ray scans for diagnosing illnesses necessitates a significant boost in efficiency, especially during crises like pandemics. The lag of 1-2 days between scans and diagnoses can critically endanger patients' lives. Consider the impact of COVID-19, with its 772 million cases and 7 million fatalities globally [1]. This pandemic triggered a 3.3 trillion dollar deficit in 2020 and a 14.7% unemployment rate in the US [2, 3]. Medical institutions grappled with surging demand, leading to lengthened wait times and exacerbated challenges, including a disturbing 117% increase in disparities in wait times [3, 4]. To overcome these obstacles, a critical reevaluation of current diagnostic approaches is imperative. Embracing innovative solutions and integrating technology are key to enhancing efficiency, minimizing delays, and optimizing healthcare outcomes.

Driven by the critical need for both speed and accuracy in medical diagnosis, researchers have turned to harnessing the transformative power of artificial intelligence (AI), particularly advanced deep learning networks. This journey began with the implementation of early but impactful classification models, such as machine learning (ML) approaches [5, 6, 7] and convolutional neural networks (CNNs) [8]. These pioneering models laid the groundwork for tackling diagnostic classification challenges, achieving promising levels of accuracy in medical image analysis.

The core of this progress lies in translating the human brain's neural architecture into mathematical models called neural networks. These networks go beyond mere computation, aiming to capture the brain's remarkable abilities to learn and generalize, as explored by Amato [9]. This initial research phase paved the way for AI applications in healthcare, especially in automating medical image analysis. However, the field has recently made rapid strides, exemplified by powerful models like EfficientNet [10]. This advanced CNN boasts a sophisticated architecture and has achieved an impressive error rate exceeding 90%. Working with 224x224x3 images, EfficientNet dynamically adjusts its width, depth, and resolution through convolutions. This showcases the field's commitment to both enhancing diagnostic accuracy and tackling the complexities of medical imaging.

AI's impact on medical diagnostics extends beyond

*Corresponding author.

Email addresses: tc922@nau.edu (Tianyi Chen), icp27@nau.edu (Ian Philippi), bqp7@nau.edu (Quoc Bao Phan), ln522@nau.edu (Linh Nguyen), carlo.cunha@nau.edu (Carlo daCunha), tuy.nguyen@nau.edu (Tuy Tan Nguyen)

CNNs. Vision transformer (ViT) models [11] offer a revolutionary approach, segmenting images and analyzing them through "transformer encoders". Unlike CNNs, ViT can distribute attention evenly across the entire image, even comparing any two segments. This unique ability lets ViT generate powerful feature representations, crucial for classifying diverse medical conditions like nuanced lung ailments. This dynamic evolution reflects the ongoing interdisciplinary effort to harness AI for complex medical challenges. Ultimately, these advancements aim to redefine healthcare practices by setting new standards for efficiency and precision in diagnosing illness.

While powerful models like EfficientNet and ViT excel in accuracy, their high computational complexity presents a challenge. Training them on new datasets demands extensive time and resources. To address this, we propose a novel fine-tuned ViT model designed to minimize training time while maintaining high accuracy. Our key contributions are:

1. We evaluate pre-trained models for the COVID-19 classification based on chest X-ray data.
2. We experimentally compare multiple models through simulations to identify the best-performing model for the classification task.
3. We strategically refine and adapt the ViT model to achieve a delicate balance of processing speed and classification accuracy.
4. We present a visual comparison of the proposed model's predictions alongside those of several recent models to highlight its enhanced effectiveness.

This paper takes a structured, three-pronged approach to exploring advanced AI models for medical image analysis. Section 2 meticulously examines existing models, unveils our groundbreaking proposition, and explains fine-tuning strategies like weight decay for optimal performance. Section 3 lays the groundwork for a fair comparison by detailing data collection, model configurations, and a thorough results analysis. Section 4 synthesizes the key takeaways, offering valuable insights into the strengths and limitations of each explored model, ultimately empowering researchers and healthcare professionals to understand how these innovative AI tools can revolutionize medical diagnostics.

2. Methodology

This section introduces the model architectures used in our experiment, as illustrated Fig. 1, and then delves into the conceptual design and fine-tuning technique of our novel ViT model, concluding with details of the chosen evaluation methods.

2.1. Proposed ViT model

The ViT model stands apart from traditional convolutional neural networks (CNNs) with its unique mathematical architecture and foundational blocks. It initiates its processing by meticulously partitioning the input image into patches of size P. These patches undergo a rigorous linear projection, followed by flattening, culminating in the formation of patch embeddings. To preserve the spatial relationships between these patches, the model dynamically generates positional embeddings, E_{pos}, during training, treating them as learned parameters. The model subsequently merges the linearly projected patch embeddings with these positional embeddings, and the resulting sum is diligently fed into the encoder blocks, preceded by a crucial normalization layer.

$$E_{patches} = [x_{class}, x_p^1 W, x_p^2 W, ..., x_p^n W] + E_{pos} \quad (1)$$

where

- W is the linear projection metrics, E_{pos} is the position embedding;
- x_{class} is class token. It is a randomly initialized array, used for storing accumulated information across the entire image and also feeding into the last layer for final classification.

The primary distinction between the ViT structure and traditional CNN architecture lies in ViT's extensive global receptive field. Unlike CNNs, which focus on local receptive fields by shifting its convolution kernels, ViT focuses on the global receptive field, which means ViT is capable of capturing relationships between distant (non-adjacent) patches within an image. This is achieved through the linear layer within the encoder block, which transforms the input into three outputs: query (Q), key (K), and value (V), by multiplying the input with three separate matrices: W_Q, W_K, and W_V. Subsequently, these outputs are processed through a similarity function, enabling each patch to learn its relation to other patches. This process, known as self-attention (SA), allows a patch to consider information from all other patches, thus capturing global dependencies. Furthermore, ViT employs multiple parallel self-attention mechanisms. These parallel operations are concatenated and then multiplied by a learnable matrix. This entire procedure is referred to as multi-headed

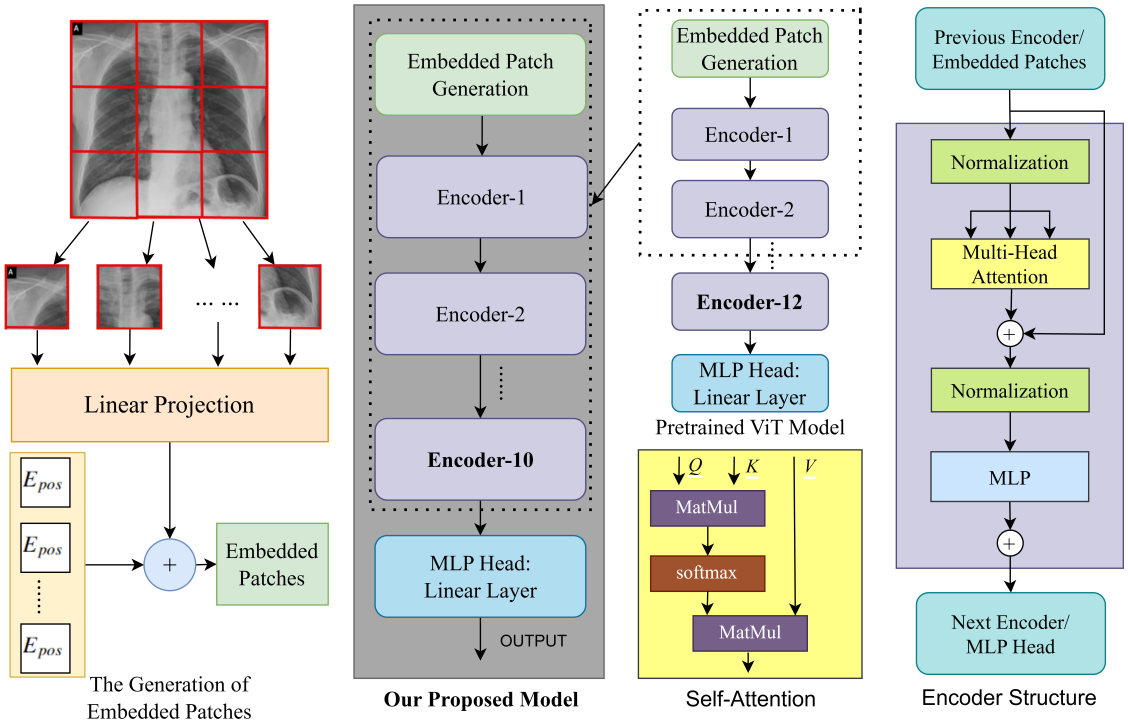

Figure 1: The structure of our proposed model.

self-attention (MSA), which enhances the model's ability to focus on various parts of the input simultaneously, thereby improving its representational power. They are realized by:

$$z_0 = E_{patches} \tag{2}$$

$$\text{Query } (Q) = z_{n-1} \cdot W_Q \tag{3}$$

$$\text{Key } (K) = z_{n-1} \cdot W_K \tag{4}$$

$$\text{Value } (V) = z_{n-1} \cdot W_V \tag{5}$$

$$\text{SA} = \text{softmax}(\frac{Q \cdot K^T}{ES}) \cdot V \tag{6}$$

$$\text{MSA} = [\text{SA}_1, \text{SA}_2, \text{SA}_3, \ldots, \text{SA}_n] \cdot W_{metrics} \tag{7}$$

where z_{n-1} is the input of the encoder block. ES stands for embedding size, which is the size of activations after patches are linearly projected. In the implemented models, its size equals patch_size × patch_size × image_channels.

At the end of MSA, a residual connection is applied:

$$z^{`} = \text{MSA} + z_{n-1} \tag{8}$$

After MSA, a multilayer perceptron (MLP) which contains two linear layers is applied to learn the local information and the complexity. It also has a residual connection at the output:

$$z_n = \text{MLP}(LN(z^{`})) \tag{9}$$

where z_n is the output of the encoder block, and LN stands for layer normalization.

The output of MLP is then fed into the next encoder block, except the last one. At the bottom, the output of MLP goes inside a linear layer (also called MLP head) in the pre-trained model, where the output of the linear layer is adjusted to the same size or number of prediction classes. The overall view of the ViT is shown in the Algorithm 1.

2.2. Weight decay

Weight decay, or $L2$ regularization, is a critical method that we use for fine-tuning our models. It adds a penalty to the loss to avoid the weights having an upheaval change during updating:

$$Loss_{updated}(\mathcal{W}) = Loss_{original}(\mathcal{W}) + \lambda \mathcal{W}^T \mathcal{W} \tag{10}$$

where $\mathcal{W}$ is trainable parameters inside the model.

Algorithm 1: Vision Transformer Workflow.

Input: Image I, E_{pos}, x_{class}, α_{mlp_ratio}, n (number of classes)
$x_{patches} \leftarrow \text{Flatten}(\text{Convolution}(I, k, s))$
$x = [x_{class}; x_{patches}] + E_{pos}$
$d = (\frac{\text{image size}}{\text{patch size}})^2$
Output: $\text{LinearLayer}(x[x_{class}], d, n)$
for $i = 0$ ***to*** *number of encoders* **do**
  $x_{norm1} \leftarrow \text{Normalize}(x)$
  MHA (Attention):
  $q, k, v \leftarrow \text{LinearProject}(x_{norm}, d, 3 \times d)$
  $x_{attn} \leftarrow \text{softmax}(q, k) \cdot v$
  $x_{atten} \leftarrow \text{LinearProject}(x_{attn}, 3 \times d, d)$
  Scaling and residual connection:
  $x' \leftarrow \text{LayerScaling}(x_{attn}) + x$
  MLP:
  $x_{norm2} \leftarrow \text{Normalize}(x')$
  $x_{fc1} \leftarrow \text{Activation}_{GELU}(\text{LinearLayer}(x_{norm2}, d, d \times \alpha_{mlp\_ratio}))$
  $x_{fc2} \leftarrow \text{LinearLayer}(x_{fc1}, d \times \alpha_{mlp\_ratio}, d)$
  $x \leftarrow \text{LayerScaling}(x_{fc2}) + x'$
return $\text{LinearLayer}(x[x_{class}], d, n)$

2.3. Experimented models

In our experiment, we employ four model architectures with pre-trained weights for transfer learning on the COVID-19 chest X-ray dataset. These include EfficientNet [12], enhanced multiscale ViT [13], EfficientViT [14], and ViT [15]. The EfficientNet, known for its computational efficiency and robustness, is constructed using a series of inverted residual blocks (MBConv blocks) [16], with varying layer configurations. These configurations are optimized through a grid search, limiting depth, width, and resolution scaling to a constant factor, alpha. The increasing popularity of transformer-based models, in time-series analysis has spurred interest in similar structures in computer vision. This trend is embodied in the ViT structure, the foundational concept discussed later. In the experiment, we implement three types of ViT structures: ViT, multiscale ViT (MViT), and EfficientViT. MViT realizes multiscale learning by learning on reduced-resolution tensors in each level of the encoder. It incorporates a pooling layer before the query, key, and value tensors. EfficientViT, similar to MViT, incorporates a multiscale learning pyramid structure through the integration of depthwise convolution, enabling the modification of tensor resolution. Additionally, EfficientViT replaces the softmax function in the similarity function with the ReLU function and employs several MBConv blocks ahead of its transformer block, significantly improving hardware computational efficiency. Each of these state-of-the-art models has demonstrated exceptional performance in classifying complex datasets, such as ImageNet-21k.

3. Evaluation Results and Comparison

3.1. Evaluation methodology

To evaluate our model's performance on the four-class classification task, we leverage key metrics such as accuracy, recall, precision, and F1 score, all of which are grounded in the fundamental concepts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). TP represents instances correctly predicted as positive, TN indicates instances correctly predicted as negative, FP signifies instances incorrectly predicted as positive, and FN represents instances incorrectly predicted as negative. These metrics collectively offer a comprehensive assessment of a classification model's performance, considering correctness, sensitivity, specificity, and the balance between precision and recall.

$$\text{Accuracy (A)} = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

$$\text{Recall (R)} = \frac{TP}{TP + FN} \tag{12}$$

$$\text{Precision (P)} = \frac{TP}{TP + FP} \tag{13}$$

$$\text{F1 Score (F1)} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{14}$$

3.2. Data collection

The COVID-19 chest X-ray dataset used in this study is sourced from the COVID-19 Radiography Database [17, 18], comprises four distinct classes. It includes 3,616 COVID-19 positive cases, 10,192 normal cases, 6,012 lung opacity (non-COVID lung infection) cases, and 1,345 viral pneumonia images, as illustrated in Fig. 2. For both training and evaluating models, the dataset is shuffled and split into 80% for training, 10% for validation, and 10% for testing.

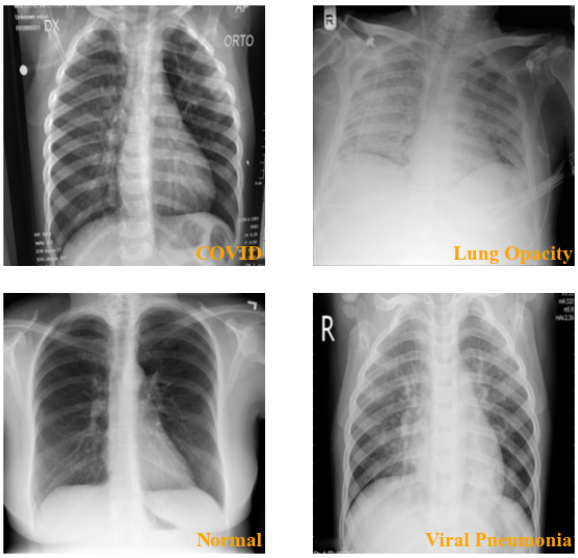

Figure 2: COVID-19 Chest X-ray Dataset.

3.3. Model configuration

We employ transfer learning with existing pre-trained models on the COVID-19 chest X-ray dataset. Following fine-tuning that involves preprocessing methods, adjustments to learning rate, batch size, and weight decay, our model is trained and tested using the settings outlined in Table 1. Additionally, the cross-entropy function served as the loss function, and AdamW and Adam are utilized as optimizers for ViT and EfficientNet models, respectively.

3.4. Result analyzing

3.4.1. Training loss analysis

To assess the training performance of the models, we conducted recognition simulations on the training and validation datasets. The results obtained from Fig. 3 indicate that the convergence capability of the models is quite effective, with the minimum value of validation loss often found in early epochs. Models based on EfficientNet and ViT show relatively promising results, with loss values typically fluctuating in the range of 0 to 0.2. However, a notable result is observed with the ViT-P16 model and our proposed model, which exhibit significantly lower validation loss. Moreover, our model demonstrates lower training loss, indicating higher accuracy achieved during the training process.

Table 1: Fine-tuning configurations.

Model	Learning Rate	Batch Size	Weight Decay
EfficientNet-B0	1.00e-03	16	1.00e-05
EfficientNet-B5	1.00e-03	16	1.00e-05
EfficientViT-B3	2.00e-06	32	1.00e-02
MViT2	2.00e-06	16	1.00e-02
ViT-Base-patch8	2.00e-06	16	1.00e-01
ViT-Base-patch16	1.00e-05	64	1.00e-02
ViT-Base-patch32	1.00e-05	64	1.00e-01
Proposed model	1.00e-05	64	5.00e-03

3.4.2. Testing performance

After the training process, we proceed with testing on the test set, evaluating the results based on the specified criteria, and finally comparing them with existing models. First, we observe that the training loss accurately reflects the achieved accuracy when the experimental models reach a high accuracy score (around 94%) as shown in Table 2 and Fig. 4. However, this is also balanced by the training time per epoch; models with higher accuracy often require more resources to train as the complexity of the model increases. For our model, which is based on ViT but with detailed and specific adjustments, we remove unnecessary layers, and retained crucial layers for feature extraction, reducing the model size. This, coupled with reduced training time (265.79 s), result in our model achieving the highest accuracy of 95.79%. On the other hand, when delving deeper into the confusion matrix of the proposed model, as depicted in Fig. 5, we can clearly discern the model's effectiveness. The most crucial class, COVID-19, achieves an accuracy of 98.58% based on $\frac{TP}{FN}$. However, the class 'lung opacity' exhibits suboptimal performance, which can be attributed to its relatively large image count and potential challenges posed by image quality in the classification task.

To provide a more comprehensive understanding of the advancements made by our proposed model, we undertook a thorough comparison of its accuracy with existing research studies, as shown in Table 3. A predominant theme in much of the existing literature has been the accurate delineation of cases into COVID-19 and non-COVID-19 categories, essentially constituting a binary classification task. Remarkably, the highest reported accuracy in these studies reached 95.11%. However, our focus on a specific class within the classification spectrum yielded exceptional results. When honing in on the task of classifying a singular class, our model demonstrated an outstanding accuracy of 99.57%. This achievement signifies the model's remarkable capability in precisely identifying instances belonging to the critical class, in this case, COVID-19.

Moreover, a comparative analysis was conducted

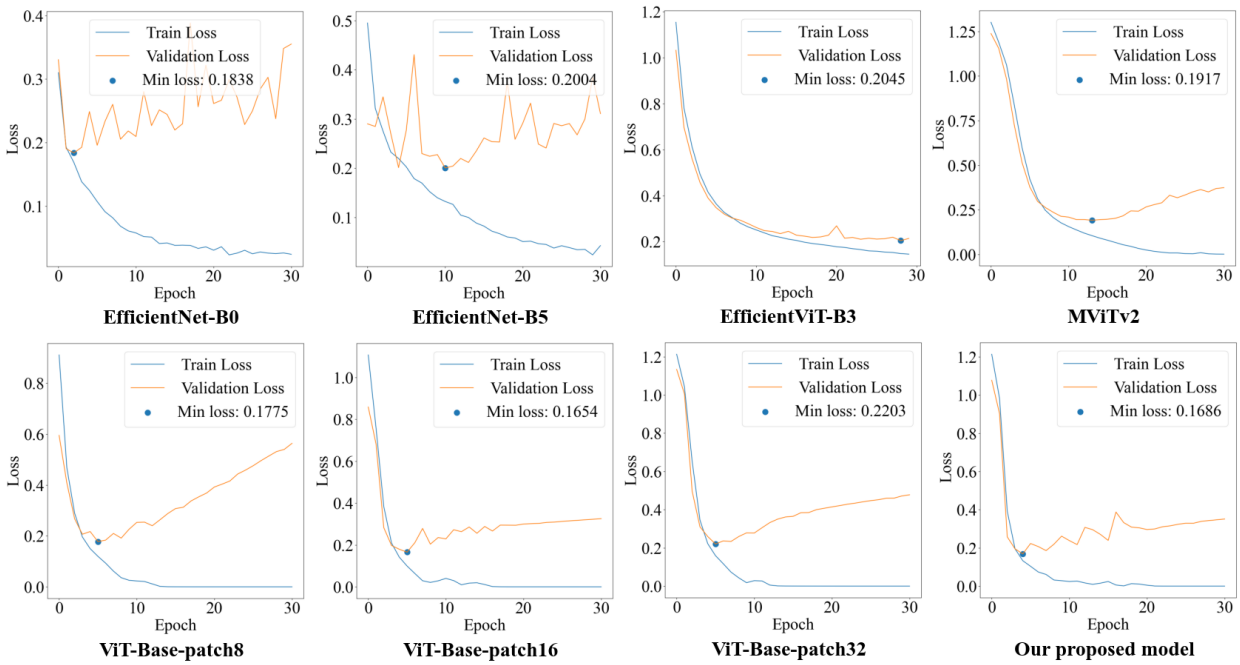

Figure 3: Train and validation losses during training process.

Table 2: Experimented models comparison.

Model	Accuracy (%)	Recall (%)	Precision (%)	F1 score (%)	Training time per epoch (s)
EfficientNet-B0	94.18	98.59	98.31	98.45	90.31
EfficientNet-B5	93.94	99.43	97.23	98.32	286.26
EfficientViT-B3	94.51	98.59	98.03	98.31	276.45
MViT2	94.41	96.88	97.71	97.29	492.49
ViT-Base-patch8	95.41	99.15	98.59	98.87	1652.94
ViT-Base-patch16	95.22	98.31	98.86	98.58	313.94
ViT-Base-patch32	93.71	98.29	97.47	97.88	89.77
Proposed model	95.79	98.58	98.87	98.73	264.79

with studies that involved the classification of multiple classes. In this scenario, our proposed model continued to exhibit noteworthy performance, achieving an accuracy of 95.79%. This result underscores the versatility and effectiveness of our model in handling more complex classification tasks encompassing multiple classes.

3.4.3. Testing visualization

In addition to evaluating the accuracy, we leverage the gradient class activation map (Grad-CAM) to visually explore our model's reasoning in classifying COVID-19 chest X-rays. This technique generates heatmaps highlighting image regions most influential for specific class predictions. Grad-CAM meticulously

Table 3: Comparison with existing studies.

Prediction Model	Classification Type	Accuracy (%)
Basic CNN [19]	Binary	93.99
Cycle GAN [20]	Binary	93.75
DenseNet201 [17]	Binary	95.11
DenseNet121 [21]	Three-Class	93.5
Resnet50-BiLSTM [22]	Three-Class	98.51
This work	Binary	99.57
	Three-Class	99.57
	Four-Class	95.79

analyzes gradients between the target class score and features extracted by a convolutional layer. Back-propagating these gradients yields neuron importance weights, which quantify each feature's contribution to the prediction. These weights then guide the creation of a weighted combination of feature maps, followed by ReLU activation to emphasize positive influences. Ultimately, the resulting Grad-CAM heatmap visually pinpoints areas crucial for the model's decision. Upon providing the COVID-19 image and its specified label, the generated Grad-CAM visualization, as shown in Fig. 6, offers valuable insights into our model's reasoning. Comparing the original image (left) with the Grad-CAM heatmap (right) reveals that the model effectively concentrates its attention within the lung region, pinpointing the area of COVID-19 infection.

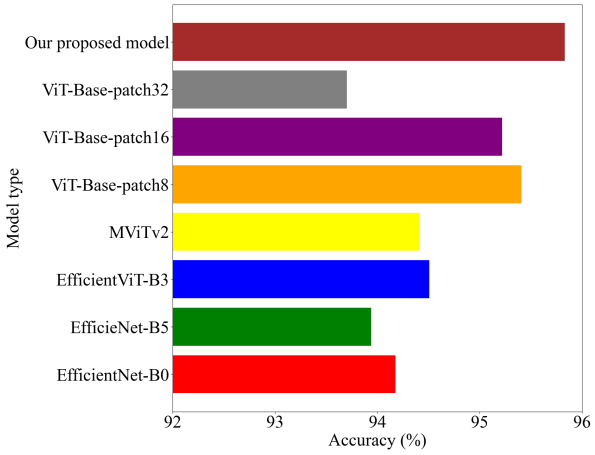

Figure 4: Experimented models performance comparison.

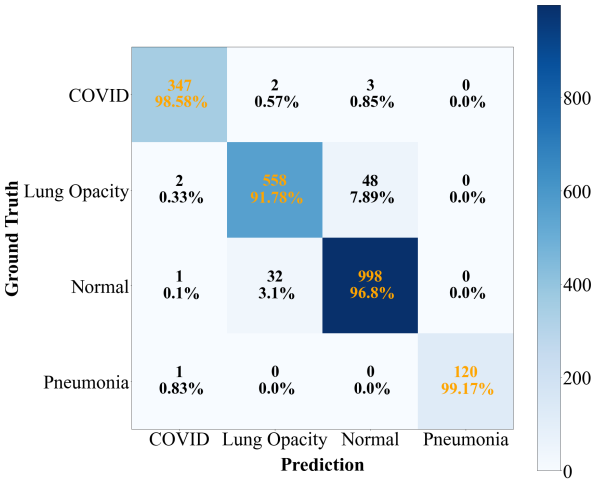

Figure 5: Confusion matrix for our proposed model.

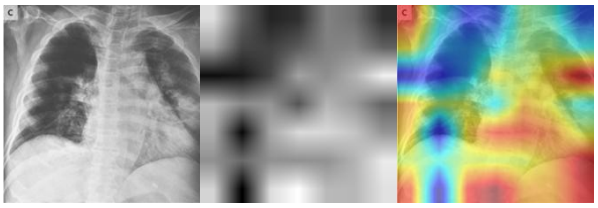

Figure 6: Grad-CAM view from our proposed model on the COVID image.

4. Conclusion

This study explores the effectiveness of AI-based models for diagnosing COVID-19 from chest X-rays. We meticulously compare diverse architectures, including EfficientNet, MViT, EfficientViT, and ViT-based models, on a carefully categorized dataset to glean crucial insights. While multiscale models exhibit tendencies towards overfitting, our proposed fine-tuned ViT model (FT-ViT) emerges as a top performer, achieving exceptional accuracy rates. Notably, it attains 95.79% accuracy in four-class classification, a striking 99.57% in a clinically relevant three-class grouping, and consistently high performance in binary scenarios. Stringent validation, employing both quantitative metrics and visualizations, solidifies FT-ViT's effectiveness. This comparative analysis across architectures highlights the superiority of our approach. In conclusion, this study not only showcases FT-ViT's potential for accurate COVID-19 diagnosis but also contributes meaningfully to medical image analysis advancements.

Acknowledgment

This work was funded under the State of Arizona Technology and Research Initiative Fund (TRIF), administered by the Arizona Board of Regents.

References

[1] W. H. Organization, Who covid-19 dashboard (Jan. 2020). URL https://covid19.who.int

[2] A. D. Kaye, C. N. Okeagu, A. D. Pham, R. A. Silva, J. J. Hurley, B. L. Arron, N. Sarfraz, H. N. Lee, G. E. Ghali, J. W. Gamble, H. Liu, R. D. Urman, E. M. Cornett, Economic impact of COVID-19 pandemic on healthcare facilities and systems: International perspectives, Best Practice Research Clinical Anaesthesiology 35 (3) (2020) 293–306. doi:10.1016/j.bpa.2020.11.009.

[3] M. Balas, D. Vasiliu, G. Austria, T. Felfeli, The impact of the COVID-19 pandemic on wait-times for ophthalmic surgery in Ontario, Canada: A population-based study 17 (2023) 1823–1831. doi:10.2147/OPTH.S409479.

[4] K. Karki, S. Priyadarshini, P. Kumar, S. Kumar, R. Kundu, K. P. Singh, A. S. Lather, K. Poonia, A. Nehra, Review on current race for Covid-19 diagnosis, Biosensors and Bioelectronics: X 16 (2024) 100432. doi:https://doi.org/10.1016/j.biosx.2023.100432.

[5] Y. Pathak, P. K. Shukla, K. V. Arya, Deep bidirectional classification model for COVID-19 disease infected patients, IEEE/ACM Transactions on Computational Biology and Bioinformatics 18 (4) (2021) 1234–1241.

[6] D. Singh, V. Chahar, V. Yadav, M. Kaur, Classification of COVID-19 patients from chest CT images using multi-objective differential evolution-based convolutional neural networks, European Journal of Clinical Microbiology & Infectious diseases 39 (July 2020).

[7] S. Tiwari, P. Chanak, S. K. Singh, A review of the machine learning algorithms for Covid-19 case analysis, IEEE Transactions on Artificial Intelligence 4 (1) (2023) 44–59. doi:10.1109/TAI.2022.3142241.

[8] Y. Pathak, P. Shukla, A. Tiwari, S. Stalin, S. Singh, P. Shukla, Deep transfer learning based classification model for COVID-19 disease, IRBM 43 (2) (2022) 87–92.

[9] A. L. F. Amato, E. M. Peña-Méndez, P. Vaňhara, A. Hampl, , J. Havel, Artificial neural networks in medical diagnosis, Journal of Applied Biomedicine 11 (2) (2013) 47–58. doi:10.2478/v10136-012-0031-x.

[10] H. Alhichri, A. S. Alswayed, Y. Bazi, N. Ammour, N. A. Alajlan, Classification of remote sensing images using EfficientNet-B3 CNN model with attention, IEEE Access 9 (2021) 14078–14094. doi:10.1109/ACCESS.2021.3051085.

[11] T.-T. Nguyen, T. V. Nguyen, M.-T. Tran, Collaborative consultation doctors model: unifying CNN and ViT for COVID-19 diagnostic, IEEE Access 11 (2023) 95346–95357. doi:10.1109/ACCESS.2023.3307014.

[12] M. Tan, Q. V. Le, EfficientNet: Rethinking model scaling for convolutional neural networks (2020). arXiv:1905.11946.

[13] Y. Li, C.-Y. Wu, H. Fan, K. Mangalam, B. Xiong, J. Malik, C. Feichtenhofer, MViTv2: Improved multiscale vision transformers for classification and detection (2022). arXiv:2112.01526.

[14] H. Cai, J. Li, M. Hu, C. Gan, S. Han, EfficientViT: Multi-scale linear attention for high-resolution dense prediction (2023). arXiv:2205.14756.

[15] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby, An image is worth 16x16 words: transformers for image recognition at scale (2021). arXiv:2010.11929.

[16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, MobileNetV2: Inverted residuals and linear bottlenecks (2019). arXiv:1801.04381.

[17] T. Rahman, A. Khandakar, Y. Qiblawey, A. Tahir, S. Kiranyaz, S. B. Abul Kashem, M. T. Islam, S. Al Maadeed, S. M. Zughaier, M. S. Khan, M. E. Chowdhury, Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images, Computers in Biology and Medicine 132 (2021) 104319. doi:https://doi.org/10.1016/j.compbiomed.2021.104319.

[18] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. A. Emadi, M. B. I. Reaz, M. T. Islam, Can AI help in screening viral and COVID-19 pneumonia?, IEEE Access 8 (2020) 132665–132676. doi:10.1109/ACCESS.2020.3010287.

[19] R. S. Saeed, B. K. Oleiwi, Deep learning model for binary classification of COVID-19 based on chest X-ray, in: 2023 15th International Conference on Developments in eSystems Engineering (DeSE), 2023, pp. 45–49. doi:10.1109/DeSE58274.2023.10099555.

[20] D. I. Morís, J. de Moura, J. Novo, M. Ortega, Cycle generative adversarial network approaches to produce novel portable chest X-rays images for Covid-19 diagnosis, in: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 1060–1064. doi:10.1109/ICASSP39728.2021.9414031.

[21] K. Rana, P. Jain, V. Shah, R. Shah, K. Ullal, M. R. Edinburgh, Detection of COVID-19 using chest X-rays, in: 2022 IEEE 7th International conference for Convergence in Technology (I2CT), 2022, pp. 1–5. doi:10.1109/I2CT54291.2022.9824121.

[22] D. E. Cahyani, D. Satyananda, M. Yasin, A. D. Hariadi, F. F. Setyawan, S. Setumin, Development of website for COVID-19 detection on chest X-ray images, in: 2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2022, pp. 330–333. doi:10.1109/ISRITI56927.2022.10053059.