International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Recognition of Covid-19 Over CT Images Using CNN and Transfer Learning

Praveen Tumuluru[1], Dr. P. Venu Madhav[2], Dr. T. Venkata Naga Jayudu[3], Konda Raveendra Kumar[4], Dr. Ravi Kanth Motupalli[5], Sunanda Nalajala[6]

Submitted: 27/01/2024 **Revised**: 05/03/2024 **Accepted**: 13/03/2024

Abstract: The emergence of COVID-19, a novel corona virus pneumonia, in 2019 has a significant impact on the growth of the global economy and the lives of individuals. Deep learning networks have been developed as a new, widely utilised image processing technique they have been employed as an unique detection approach in clinical practise to retrieve medical info from CT pictures. Yet, because to the medical features of Corona virus Computed tomography scans. As a result, Using the current deep learning method, diagnosis is tough. The global feature extraction advantage of the Transformer module and CNNs module is provided, and the concurrent trans model (TransCNN Net) based on Transformer is employed to fully use the local extraction of features capabilities of the CNN Model.. This is done in line with the COVID-19's medical characteristics on the CT scans. By extracting features from two branches and fusing them in opposite directions, the cross-fusion notion results in a bi-directional feature fusion structure. .A feature fusion module then joins the parallel branch structures to create a model that can identify scale-dependent properties. The classification accuracy of covid data is 96.7, which is greater than Transform Network (Diet-B). The outcomes show increased accuracy. A new approach for diagnosing COVID-19 is also provided by this model, and by it encourages the growth of instantaneous identification pulmonary issues brought on by a corona infection by integrating machine learning with medical imaging. This approach enables an accurate and speedy diagnosis, saving precious lives.

Keywords: *Deep Learning, Transformer, Convolutional neural network, Computed Tomography, and Corona.*

1. Introduction

The corona virus become widespread throughout the globe, and clinical verification of the COVID-19 diagnosis using lung-specific CT scans has been made. Recent research [1,2] has shown that the diagnosis of corona can benefit greatly from the knowledge obtained from lung-related CT scans. However even if there are some autonomous and intelligent systems, their methods remain with a huge space for development and depend highly on human judgment and heavily in expertise of medical professionals. The diagnosis of COVID-19 must consider both the local and general characteristics of the lesions because it is not possible to make the diagnosis just based on geographic aspects. A doctor's professional knowledge and relevant experience are necessary for the examination and diagnosis of CT images, which is a very challenging process[30]. For the doctor to perform manual treatments, it requires a lot of effort and time. The morphology of some COVID-19 CT images also resembles that of traditional imaging techniques for pneumonia. The outcomes of computed tomography scans demonstrate three unique infection types: (A) new coronavirus resulting from corona disease, (B) ordinary infection, and (C) control subjects. The examples of data are displayed in Fig. 1. The primary characteristics of pneumonia induced by the novel coronavirus include GGO, that commonly develop over both sides and in the surrounding area; The paving stone sign, thicker interstitial membrane, and interstitial line merging with GGO, sometimes known as "crazy paving," may occasionally appear as the condition worsens. Most people think that this type of rash appears late in the disease' development. Some people will have signs like local Hemangi ectasia.

The following picture, which is presented in Fig. 2, shows the primary physical feature of new coronary pneumonia. The CT pictures of COVID-19 show both local and global

1 Koneru Lakshmaiah Education Foundation - INDIA
ORCID ID : 0000-0001-9426-205X
2 P.V.P Siddhartha Institute Of Technology - INDIA
ORCID ID : 0000-0003-1349-4014
3 Srinivasa Ramanujan Institute of Technology - INDIA
ORCID ID : 0000-0003-3864-2351
4Guru Nanak Institutions Technical Campus– INDIA
ORCID ID : 0009-0000-2112-879X
5Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering & Technology-INDIA
ORCID ID : 0000-0003-4893-5166
6Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering & Technology-INDIA
ORCID ID : 0000-0001-9919-4180
ORCID ID : 0000-0001-9919-4180
** Corresponding Author Email: praveenlurur@email.com*

anomalies, including such large-area GGO. Local Hemangi ectasia and crazy paving are two examples of the region's unique traits. It is distinguished by the fusion of regional and global elements[31,32].

with various scales. This technique still only uses local features in its basic form because of t the neighborhood of convolution kernels' receptive fields. Using a cascaded branch network is a different strategy [4].

In the end, The process of feature fusion is straightforward to understand. after different branches extract features on various scales. Convolution kernel size must be increased or dilated convolution must be used[5]. The convolution kernel's excessive size is the issue, which will limit the network's capacity for generalization [6]. utilizing the Liu Z et al. [7] image pyramid technique. It has been determined how to extract visual features at various scales. In order to create picture pyramids, the photos are scaled at various ratios[9].

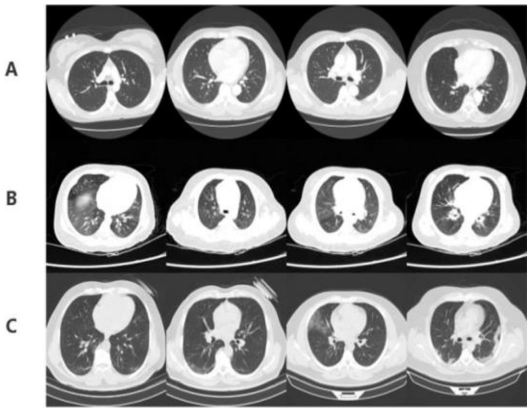
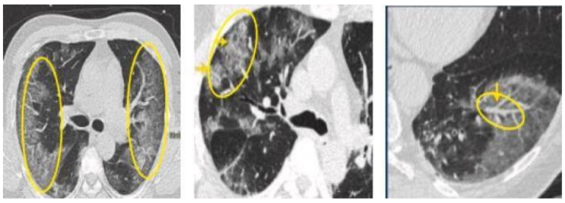


Fig. 1. Example of data set.



(a)GGO (b) crazy paving (c) hemangiectasis

Fig. 2. Biomedical characteristics of covid CT images**.**

Next, feature maps are created by extracting features at various ratios from each layer of images. The pyramid of a picture is a group of pictures created based on a single original image and organized in a pyramidal structure with increasingly decreasing quality. It is achieved by gradual down sampling, which lasts until a certain termination condition is satisfied. A pyramid-like structure is used to compare the image. The resolution and size of the image decrease as the level rises[11].

While features of various scales can accurately and richly convey semantic information, the processing speed will be slowed, and the amount of data generated by overlapping pairs of these characteristics will rise. The FPN suggested by Lin T Y et al [8]. The suggested approach for feature fusion includes several resolutions, That is, feature enhancement at various levels is achieved by adding feature maps of various resolutions and subsampling low-resolution features element-by-element [29, 30]. This method's efficiency is improved and the amount of additional computation is reduced because it solely uses network-based cross-layer connections and element-wise addition[12].

Significant advancements in computer vision have been made by Vision Transformer (ViT) [13] and Deit [14], which show the benefits of global processing and significantly outperform CNN. The network's parameters will significantly rise if the sole structure utilized to extract characteristics is the Transformer structure[10,21]. The advantages of CNN's deep convolutional and point multiplication convolution in evaluating local characteristics determine whether CNN [16, 15, 17] and its related enhancement work [19,18] will continue to occupy a prominent place in the field with very little computation. CNN was included to the visual Transformer by Wu Haiping et al. [20].

Four-layer convolution came first, then numerous Transformer modules, and eventually classification. CNN and Transformer both have a similar serial format [22,28]. The Transformer network's branches are chosen to extract the global features of pictures by leveraging the network's global receptive field qualities, in contrast to the aforementioned traditional approaches, which alter the convolution kernel's size to extract the features of various scales. This study makes following benefaction [25, 26].

1) A CNN-based and Transformer-based branch network model is suggested.

2) Build a framework that can merge characteristics from two branches in a bi-directional manner.

3) Inventively apply the model proposed in this study on large COVID-19 dataset COVIDx- CT;

4) Be doing a component visualizing test to show the model's logic and verify the validity of the classification base [33].

2. Methodology

Convolutional neural networks have demonstrated significant improvements in retrieving picture features. Because of the way CNN is built, It excels in capturing local elements from images that are easily discernible. CNN's audience in its local market, a network topology made up of a convolution kernel, is its most notable characteristic. Yet, Visible localized lesion characteristics and dispersed all features are among the medical criteria of processed picture, This is a CT examination performed to identify COVID-19. So because data set is distinct, this article uses the Transformers component to extract the feature map from CT scans and combine them to enhance classification results.

Convolutional Neural Network and Transformer modules are used. a classification model is presented in this paper for identifying and extracting the characteristics of the use of 3 Covid computed tomography images is suggested. Precise instructions are listed below.

1) The corona virus CT-scan image is processed by Transformers branch module, and the image's features are retrieved as a consequence of a picture's global receptive field characteristics.

2) Local COVID-19 CT image features are obtained via convolution convolutional networks field features.

3) To integrate the characteristics of the two branches, a dual-direction function matching structure is constructed Within the two branches. Bi-directional pattern fusion is employed in this case to provide richer, more complete features and improve classification accuracy.

Transformer encoder:

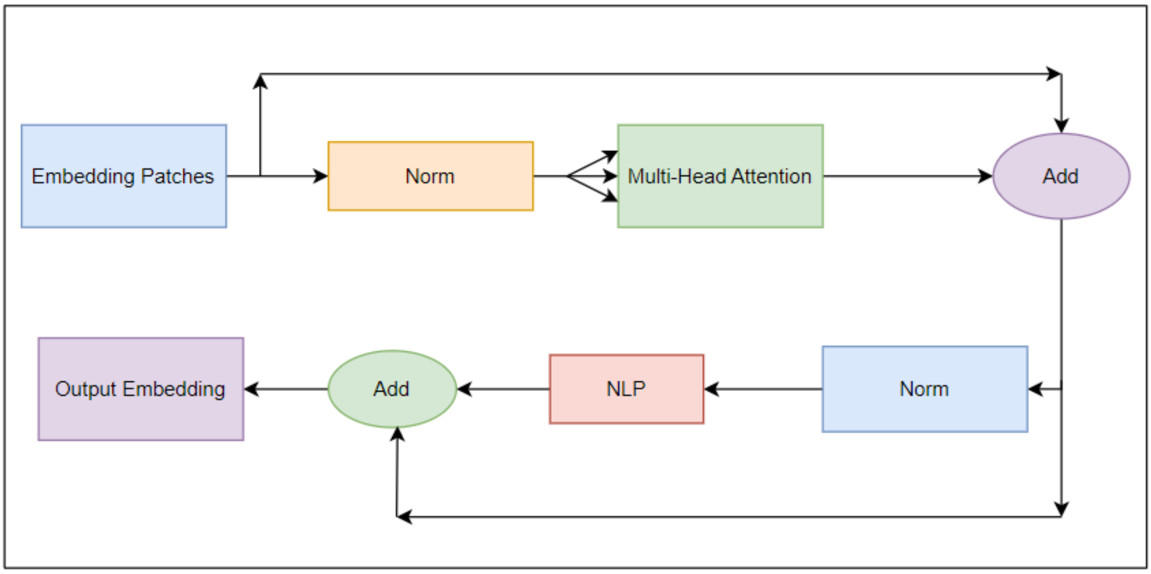


Fig. 3. Encoder model in Transformer

Scaled Dot-Product Attention: $a_{1,\ i=q}^{1} * k^{i}$ /root (d)

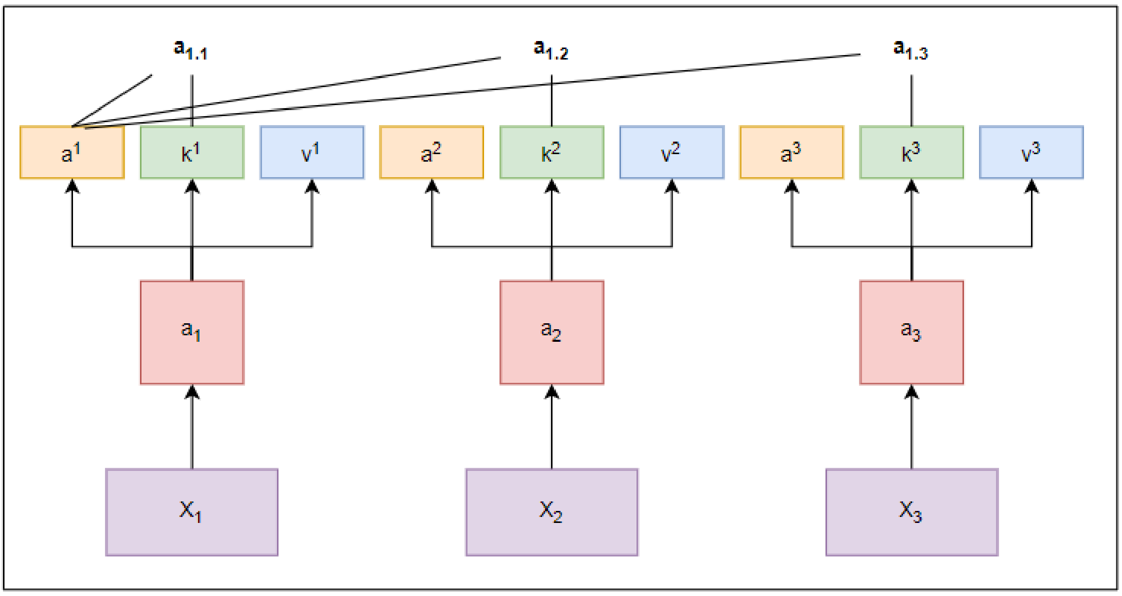


Fig. 4. Self-Attention module.

Google's model, Transformer [23], was introduced in 2017 and is applied area of NLP. A key component of the global receptive field, which is significant in transformer network structure, is attention processes [27]. In another light, "Transformers" could be seen as a unique CNN with a universal sensation field. Encoder and decoder make up the basic construction of a transformer, which is made up of two sections.

The structure of this branch's transformer, which only uses the encoder portion, is depicted in Fig.

3. Transformer uses a self-attentive module, a normalized dot product attention mechanism. The input xi feature vector is represented by ai, as shown in Fig. which is then matched to the equivalent values of q_i, k_i, and v_i. To multiply dots, or The Self Attention mechanism is in charge of carrying out matrix operations on all q and all k. By normalizing each input mapping, the attention weight matrix must first be obtained. The Transformer design has been utilized in the subject of natural language processing ever since inception, terms like query, key, and value have been in use.

2.1. Transformer module:

Table 1: Network characteristics of the upgraded VGG-19

Module	Convolution Kernel	Output
Inputs	224*224*3	
Block1	2*conv 3*3*64 pool 2*2	112*112*64
Block2	2*conv 3*3*64 pool 2*2	56*56*128
Block3	4*conv 3*3*256 pool 2*2	28*28*256
Block4	4*conv 3*3*512 pool 2*2	14*14*512
Block5	4*conv 3*3*512 pool 2*2	7*7*512
FC_1	1*1*4096	4096
FC_2	1*1*1024	1024
FC_3	1*1*256	256

When the query's entry dimensions are d_k for each key and $d_{v,}$ for the value dimension, the action of point multiplication of The weight is determined for each key, followed by a division by the root of d k. is then determined using the softmax function.

$$\text{Attention}(Q, K_i, V_i) = softemax(\frac{Q^T K_i}{\sqrt{d_k}})V_i$$

(1)

The following is the output matrix. The terms "query, keys, and values" (or "Q, K, and V") are used in the context of computer vision.

$$\text{Attention}(Q, K, V) = softemax(\frac{Q^T K}{\sqrt{d_k}})V \qquad (2)$$

The calculation formula is as describes, and There are two Norm and Add layers in the encoder.

LayerNorm(X+MultiHeadAttention(X)) (3)

LayerNorm(X+FeedForward(X)) (4)

Both Multi-Head Attention(X) and Feed Forward(X) represent the output. The input and output dimensions can be added because they are the same. X is a graphical representation of the Multi-Head Attention or Feed Forward input. The Feed Forward layer is joined by two entirely interconnected layers. The first layer's activation function is relu, not the second layer's activation function. As a result, the following formula is used.

$$\text{Max}\ (0, XW_1 + b_1)W_2 + b_2 \qquad (5)$$

The dimensions of the matrix's output produced by feed forward models are consistent with the input, X, which is the input.

2.2. CNN (Convolution neural network) module:

To maintain the same domain and increase the network's depth, The VGG 19 model substitutes several 3*3 convolution kernels for the larger convolutional network that AlexNet utilized.

Moreover, A 3 * 3 convolution kernel can be used in place of a large convolution kernel to reduce the parameters. The VGG 19 network performs remarkably well when it comes to picture categorization. Five modules together known as "convolution pool modules" make up the network's simple structure. where "pool layer" describes the convolution kernel layer and "convolution" is used to describe the convolution kernel. of the convolution. Five convolution pool modules are connected after them, and the final layer has 256 neurons. (See Table 1).

2.3. Bi-directional fusion module:

The CNN branch and Transformer branch are fused in both directions via the fusion layer. Refer to the bidirectional bridge construction [24], Most crucially, the ability to exchange local and global information between the Transformer branch and Convolution CNN is realized by adopting a straightforward bi-directional fusion structure, this successfully blends the effectiveness and efficiency of the part- and whole-model approaches. Thereafter, several features are fused by the pool block of Block 2 is the first Transformer of VGG 19 network's encoder block. We chose to link at these two places since, if the connections were made further from, the difference in processing rates between the two branches would slow down parallel computation.

For this reason, the first half of the two branches was combined. As convolution layers are layered, Local characteristics ultimately disappear as the receptive field expands, but the global features gained in the first half of Transformer are more complete. Global features and layer upon layer superposition will be lost as a result of the attention mechanism operating continuously. Conjoin CNN and Transformer by parallelizing them and connecting them using the self-attention structure in CNN and Transformer (show in Fig. 5). To extract local features, CNN uses the image as its input.

Transformer and Convolutional Neural Network are connected by a bi-directional feature fusion architecture it combines global and local properties in counterintuitive directions. The bidirectional feature fusion structure's concrete framework will next be discussed.

In Fig. 5, The proposed technique combines regional characteristics with global characteristics. in a unidirectional CNN Transformer structure .

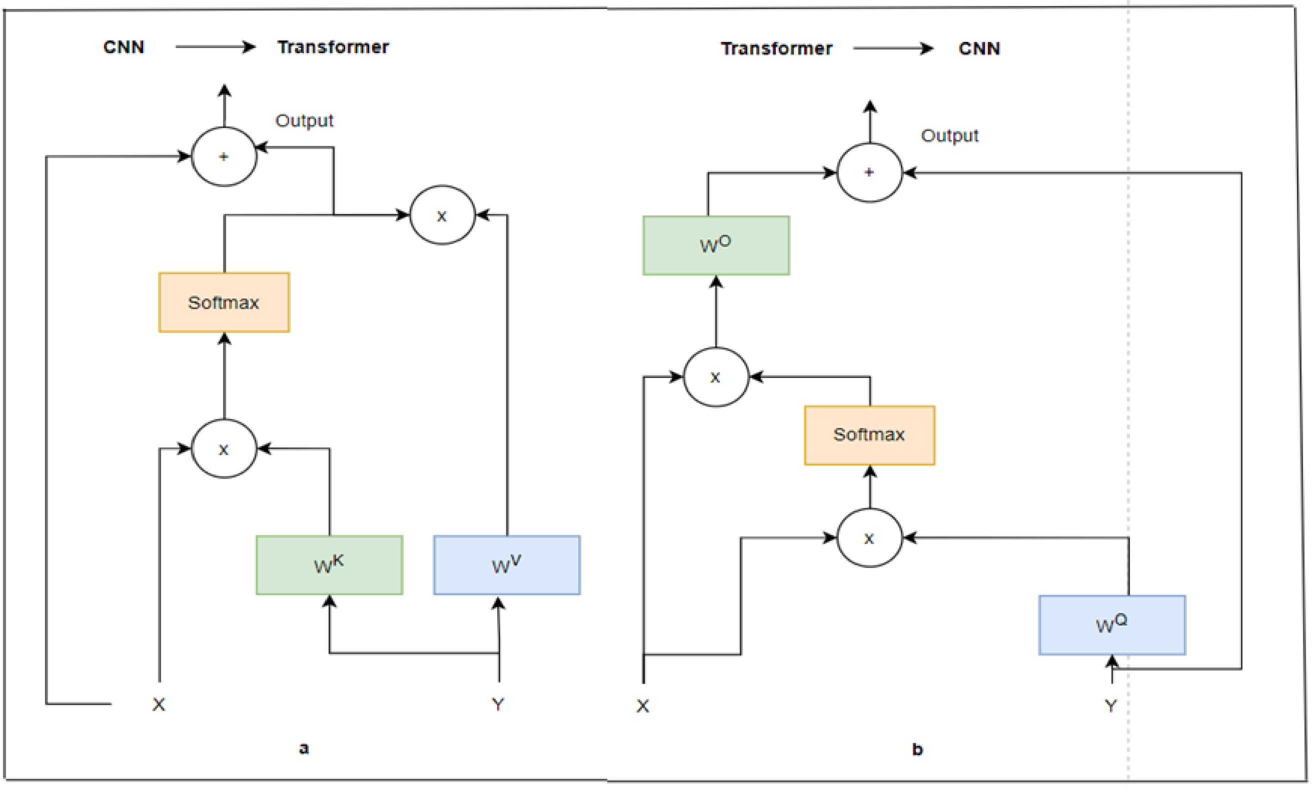


Fig. 5. Two-way feature fusion structure.

Below is a definition of local to global fusion.

$$\text{head}_i = \text{Attention}\ (Y_h\ W_h^Q\ ,\ X_h,\ X_h) \qquad (6)$$

$$Y^{out}\ =\ Y + [\text{Concat}(\text{head}_1,\text{head}_2,\ \ldots,\ \text{head}_h)]W^o \qquad (7)$$

The formula specifies that WhQ is the projection matrix for the query and that Attention (Q, K, V) is the standard. Attention function on V, K, and Q. Wo combines a lot of headers (2).

Q serves as the global input feature Y is represented by K and V and acts as the local input feature X.

To the global feature Y, W_h^Q and W^o are applied.

Likewise, formula (8) and is used to determine how to calculate from global to local feature fusion structure

head = Attention (Xh, Yh Whk, YhWhv) (8)

$$X^{out}\ =\ X\ +\ [\text{Concat}(\text{head}_1,\text{head}_2,\ \ldots,\ \text{head}_h\ )] \quad (9)$$

In which the projection matrices of keys and values W_h^K and W_h^V.

A key and a value make up the global feature Y, whereas a query makes up the local feature X. On This feature fusion's schematic diagram is shown in the right figure of Fig. 5 in this way

(Transformer CNN). output and input the global mark Y belongs RM*d, while the local feature graph X belongs Rhw*C, It consists of the spatial coordinates hw (hw = h*w, where h and w are the height and width of the feature graph, respectively) and the C channels. the size and quantity of feature blocks, M and d, respectively, are the two inputs needed by the CNN-Transformer block.

2.4. The network architecture and the whole model design:

In this work, a parallel mode merging CNN and Transformer is demonstrated after converting the typical serial structure into a parallel structure. introduced. As seen in Fig. 6, Transformer's inputs are the vector of the entire image, and the number of parameters is the global receptive field, allowing one to utilize the benefits of Transformer while extracting global features. CNN creates a trans parallel network topology and parallelizes the roll integrating branch and the Transformer branch in order to effectively extract local information simultaneously. Fig. 6 shows how the bi-branch parallel structure's central region is changed into a sub feature fusion structure

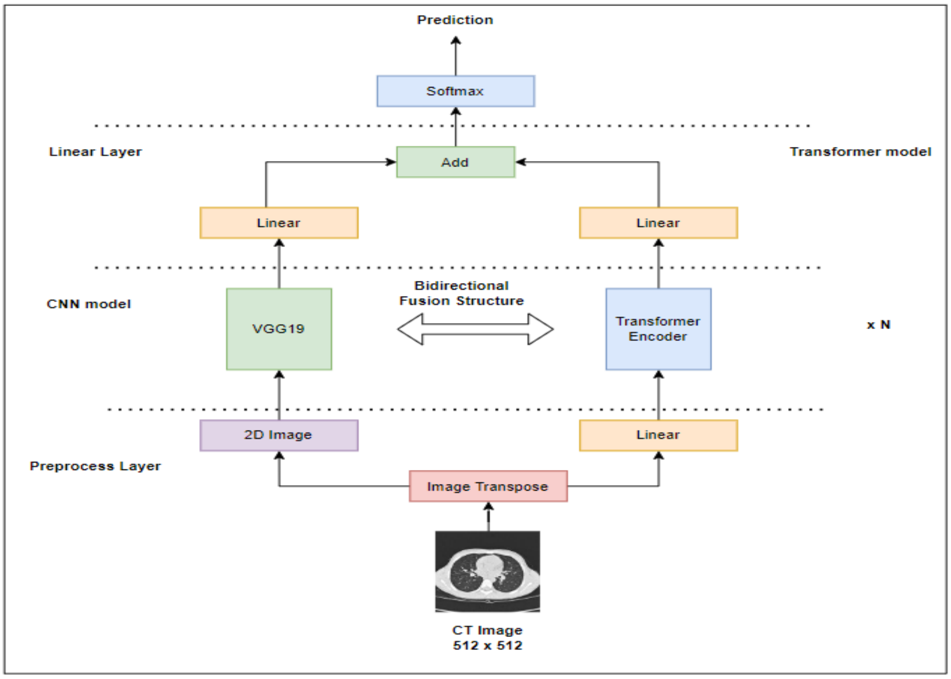


Fig. 6. Overall structure of the model

3. Experiment and Result Analysis

3.1. The setting and equipment for the experiment:

In this study, the technique is implemented using the 64-bit Ubuntu-18.04.1 operating system.

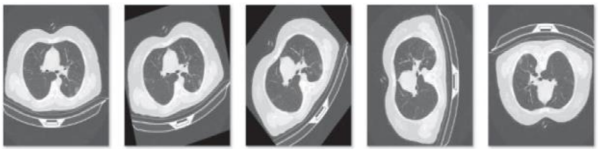
The 3-card parallel training was carried out using an Intel(R) Xeon(R) CPU E 5-2695 and a high-performance NVIDIA V GPU, with each graphics card running on a system with such a storage space of 16 gigabyte and 12 gigabytes of RAM. CUDA 11.0, CUDNN 7.6, and Pytorch 1.7.0 were used to create and train the model. And with learning basics rate set to 0.0005, pre-training was done on the ImageNet dataset. The random inertia gradient descent method is used in the optimization process with a batch size of 128.The system had stabilized in the 48th epoch.

3.2. Dataset:

This study introduces COVID-19-CT, a benchmark CT image data set made up of 194,922 images of 3745 patients ranging in age from It has been firmly clinically established that the range is 0 to 93 years (median age 51 years).

The data collection was created using CT imaging information that was gathered by the China National Bioinformatics Centre.

In the corona – CT benchmark data set, chest CT volume results in three distinct infection types: the common cold, common pneumonia, and normal control. how many people experience three infections on average.



(a)CT-0° (b)CT-15° (c)CT-45° (d)CT-90° (e)CT-180°

Fig. 7. Example diagram of data set pre-processing

Table 2 displays the many types of training, verification, and testing. By applying data augmentation, the sample size is raised, and data enhanced by shifting at various angles. (180, 90, 15 and 45). Fig. 7 shows the improved data, which has been normalized. The data set used to train the Transformer structure model needed to be expanded in order to get better results.

Table 2: Data Partition

Type	Normal	Pneumonia	COVID 19	Total
Train	35,900	25,400	82,000	1,43,700
Val	11,800	7400	6200	25,000
Test	12,200	7300	6000	25,000

3.3. Experimental result:

As depicted in Figs. 8, the training loss and precision convergence process.

The model has strong feature extraction skills and a speedy convergence rate when being trained because it has previously been trained on the ImageNet dataset.

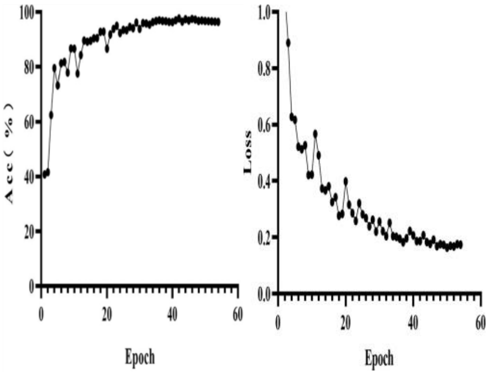
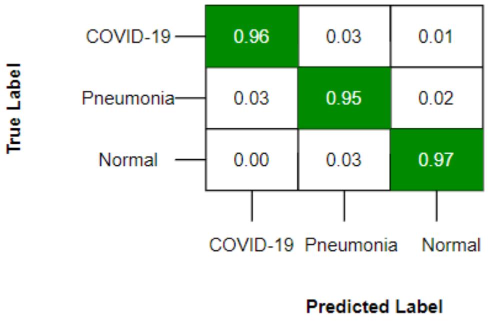


Fig. 8. Network's training accuracy and Network training lost.



True Label	COVID-19	Pneumonia	Normal
COVID-19	0.96	0.03	0.01
Pneumonia	0.03	0.95	0.02
Normal	0.00	0.03	0.97

Fig. 9. Using the Trans-CNN Network test, a confusion matrix was discovered**.**

According to the confusion matrix of the model, which is depicted in Fig. 9, The model's classification accuracy for common pneumonia after loading the trained model is approximately 95%, Almost 96% of COVID-19 was correctly identified using this method, and 96.9% of the time.

3.4. Experimental evaluation index:

As part of our test, five metrics—specificity, precision, sensitivity, accuracy, and F 1 score—were utilized to assess how well our suggested models performed. The calculation of the aforementioned indices includes four primary factors: true positive, false positive, true negative and false negative. TP represents the proportion of COVID-19caused pneumonia that was correctly classified; TN the proportion of COVID-19-caused normal that was correctly classified; FP the proportion of COVID-19-caused pneumonia that was incorrectly classified as normal; and FN the proportion of normal that was incorrectly classified.

Specificity measures how well our models did at distinguishing the test set's normal images. It is passable.

$$\text{Specificity} = TN/(TN+FP)$$

Sensitivity is defined as the percentage of genuine positives (pneumonia verified by COVID-19) that are accurately detected.

$$\text{Sensitivity} = TP/(TP+FN)$$

Accuracy is a measure of how well our model can categories data, and is expressed as follows:

$$\text{Accuracy} = (TP + TN)/(TP+FP +FN+TN)$$

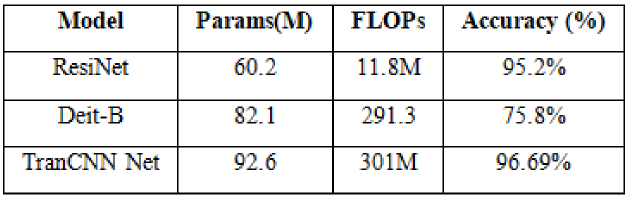
According to precision's definition:

$$\text{Precision} = TP/(FP+TP)$$

The F1 score evaluates your classification skills: F1 score = 2 × Precision × Sensitivity/Precision+Sensitivity

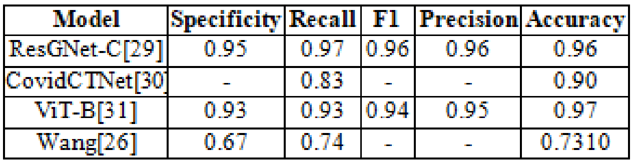
On this collection of data, the pure Transformer network Deit's classification accuracy is 95.2% and 75.8%, respectively. The standard Resnet-152 network, which has a lower classification accuracy, has a lesser classification effect on COVIDx-CT than this network does (See Table 3).

Table 3: Several classification models are compared to Tran-CNN Net.



Model	Params(M)	FLOPs	Accuracy (%)
ResiNet	60.2	11.8M	95.2%
Deit-B	82.1	291.3	75.8%
TranCNN Net	92.6	301M	96.69%

Table 4: comparison with suggested approaches.



Model	Specificity	Recall	F1	Precision	Accuracy
ResGNet-C[29]	0.95	0.97	0.96	0.96	0.96
CovidCTNet[30	-	0.83	-	-	0.90
ViT-B[31]	0.93	0.93	0.94	0.95	0.97
Wang[26]	0.67	0.74	-	-	0.7310

Our suggested method is contrasted with a few other methods that are currently in use to further verify it. Results specifically, recall, F1, precision, and accuracy are where the model provided in this study excels, as demonstrated in Table 4.

3.5. Visualization of classification basis:

The important component model categorization areas are visualized using Grad-weighted class activation mapping (Grad-CAM), and the outcomes are displayed in Fig. 10 to better assess the model's validity.

4. Discussion

In order to recognize CT images, we developed and applied an architecture for CNN and Transformer-based multi-branch capability. The experimental findings demonstrate the value of this approach in identifying COVID-19 in CT scans. The key benefit of this approach is, it develops a bi-branch network while fully utilizing The ability of CNN and Transformer to extract characteristics at multiple scales. The outcome is superior to one Transformer network and one CNN structure network.When building a CNN- and Transformer-based bi-branch feature fusion framework Our study examines the performance of several CNN and Transformer network combinations while constructing a bi-branch network, and it finds that the classic The ideal combination is between Transformer network and VGG-19. (Fig. 11). An image's general characteristics are referred to as global features. Features like colour, texture, shape, and others are frequently used global features. We developed tests to get an activation map of Transformer at various phases in order to investigate the Transformer's diminishing effect on extracting global characteristics. The global features in Columns c, the feature map produced by Transformer Block 3, and Columns d, the output character diagram of Transformer Block 5's network, are significantly reduced by Transformer's self-attention approach. We see that it exhibits more concentrated local features after going through the final block of the transformer. As anticipated, the global properties gradually decline as the number of Transformer layers rises.

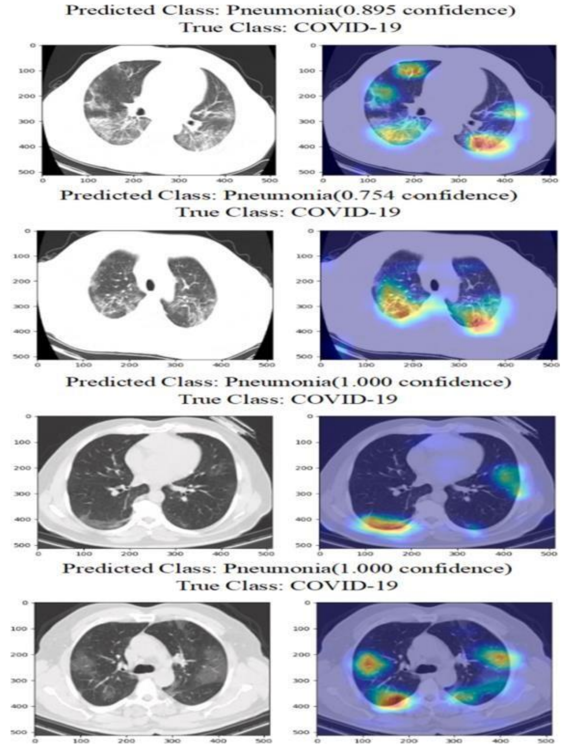

Predicted Class: Pneumonia(0.895 confidence)
True Class: COVID-19
Predicted Class: Pneumonia(0.754 confidence)
True Class: COVID-19
Predicted Class: Pneumonia(1.000 confidence)
True Class: COVID-19
Predicted Class: Pneumonia(1.000 confidence)
True Class: COVID-19

Fig. 10. Comparison diagram of feature visualization.

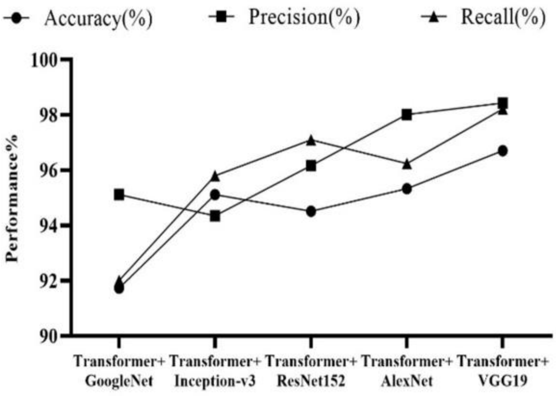


Fig. 11. Analyzing the interactions between various CNN topologies and Transformer combinations.

To lessen the loss of crucial feature information, use a feature fusion structure. We have frequently tested when there are many fused design characteristics in two branch nodes.

5. Conclusion

Based on transformer and convolutional neural networks, this paper recommends a bi-branch feature fusion network structure. The benefits of the two branches' respective local and global features are used to extract the CT scan features. Similar to this, the fusion layer bidirectionally combines the characteristics of the two branches to analyze the network's data in parallel, speed up network operation, and generate better classification results.

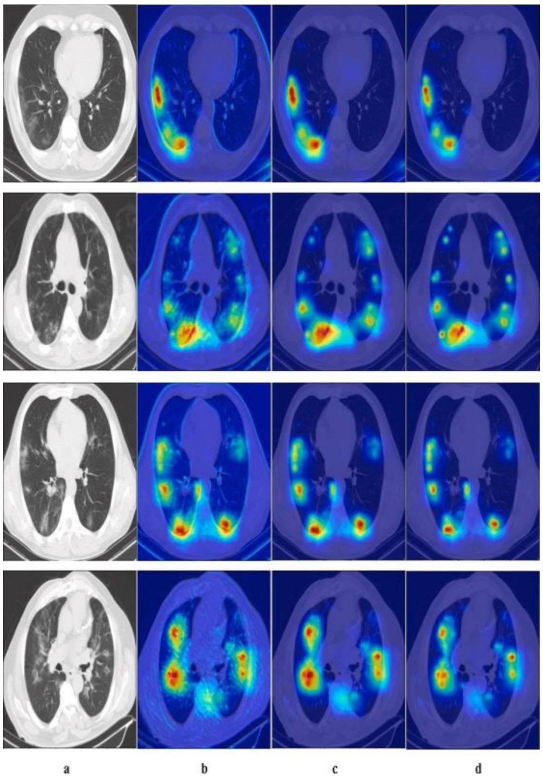


Fig. 12. Transformer takes the global feature impact and extracts it.

The approach can extract features with both local and global data and has a straightforward structure and good generalization capabilities. Effectiveness is high. The classification accuracy for the COVID-19 challenge utilizing CT scans is 96.7%. In conclusion, the classification of medical images will greatly benefit from our research. The categorization of CT pictures used in this experiment, however, is based on a large number of patient CT scans, this is less feature-rich and suffers from the problem of inaccurate patient diagnostic data.

So, with numerous CT images of the same patient, it is possible to investigate three-dimensional reconstruction. based on perfecting the network and explore this further in subsequent work. The original picture is shown in column as of Figure 12 and the feature map created by Transformer's Block 1 is shown in column b. It is obvious that Block 1 of Transformer has adverse global properties.

References:

[1] Z. Ye, Y. Zhang, Y.i. Wang, Z. Huang, B. Song, Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review[J], Eur. Radiol. 30 (8) (2020) 4381–4389

[2] A.J. Rodriguez-Morales, J.A. Cardona-Ospina, E. Guti´errez-Ocampo, R. VillamizarPena, Travel Med. Infect. Dis. 34 (2020) 101623.

[3] Z. Lin, Z. Luo, L. Zhao, et al., Multi-scale convolution target detection algorithm with feature pyramid[J], J. ZheJiang Univ. (Eng. Sci.) 53 (3) (2019) 533–540.

[4] Cheng Weiyue, Zhang Xueqin, Lin Kezheng, et al. Deep Convolutional Neural Network Algorithm with Fusing Global and Local Features. [J/OL]. Journal of Frontiers of Computer Science and Technology: 1-11 [2021-09-02].

[5] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions [J]. arXiv preprint arXiv:1511.071 22, 2015.

[6] H.T. Cheng, L. Koc, J. Harmsen, et al., Wide & deep learning for recommender systems[C], in: Proceedings of the 1st workshop on deep learning for recommender systems, 2016, pp. 7-10.

[7] B. Singh, L.S. Davis, An analysis of scale invariance in object detection snip[C], Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2018:) 3578–3587.

[8] T.Y. Lin, P. Dollar, ´ R. Girshick, et al., Feature pyramid networks for object detection[C], Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017:) 2117–2125.

[9] S.-H. Wang, M. Attique Khan, V. Govindaraj, S. L. Fernandes, Z. Zhu, Y.-D. Zhang, Deep rank-based average pooling network for COVID-19 recognition[J], Comput. Mater. Continua 70 (2) (2022) 2797–2813.

[10] S.H. Wang, X. Zhang, Y.D. Zhang, DSSAE: Deep stacked sparse auto encoder analytical model for COVID19 diagnosis by fractional Fourier entropy[J], ACM Trans. Manage. Inform. Syst. (TMIS) 13 (1) (2021) 1– 20.

[11] S.-H. Wang, Z. Zhu, Y.-D. Zhang, Patch Shuffle convolutional neural network for COVID-19 explainable diagnosis[J], Front. Public Health 9 (2021).

[12] Z. Huang, X. Liu, R. Wang, M. Zhang, X. Zeng, J. Liu, Y. Yang, X. Liu, H. Zheng, D. Liang, Z. Hu, FaNet: fast assessment network for the novel coronavirus (COVID19) pneumonia based on 3D CT imaging and clinical symptoms[J], Appl. Intell. 51 (5) (2021) 2838–2849.

[13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., An image is worth 16x16 words: transformers for image

recognition at scale[J]. arXiv preprint arXiv:2010. 11929, 2020.

[14] H. Touvron, M. Cord, M. Douze, et al., Training data-efficient image transformers & distillation through attention[C], In: International Conference on Machine Learning. PMLR, 2021, 10347-10357.

[15] A. Howard, M. Sandler, G. Chu, et al., Searching for mob-ilenetv3[C], Proceedings of the IEEE/CVF International Conference on Computer Vision. (2019) 1314–1324.

[16] A.G. Howard, Z.u. Menglong, C. Bo, et al., Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv, 2017.

[17] M. Sandler, A. Howard, Zhu Menglong, et al., Mobilenetv2: Inverted residuals and linear bottlenecks [C], In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018:4510-4520.

[18] K. Han, Y. Wang, Q. Tian, et al., Ghostnet: More features from cheap operations[C], Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. (2020) 1580–1589.

[19] Li Yunsheng, Chen Yinpeng, Dai Xiyang, et al. MicroNet: Improving Image Recognition with Extremely Low FLOPs[J]. arXiv preprint arXiv:2108. 05894, 2021.

[20] Wu Haiping, Xiao Bin, Codella N, et al. Cvt: Introducing convolutions to vision transformers[J]. arXiv preprint arXiv:2103.15808, 2021.

[21] B. Graham, A. El-Nouby, H. Touvron, et al., LeViT: a Vision Transformer in ConvNet's Clothing for Faster Inference[J]. arXiv preprint arXiv:2104.01136, 2021.

[22] T. Xiao, M. Singh, E. Mintun, et al., Early convolutions help transformers see better [J]. arXiv preprint arXiv:2106.14881, 2021.

[23] A. Vaswani, N. Shazeer, N. Parmar, et al., Attention is all you need[C], Advances in neural information processing systems, 2017, p. 5998-6008.

[24] Y. Chen, X. Dai, D. Chen, et al., Mobile‘-Former: Bridging MobileNet and Transformer[J]. arXiv preprint arXiv: 2108. 05895, 20.

[25] S. Lu, Z. Lu, J. Yang, M. Yang, S. Wang, A pathological brain detection system based on kernel based ELM[J], Multimedia Tools Appl. 77 (3) (2018) 3715–3728.

[26] S. Wang, B.o. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, B.o. Xu, A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19) [J], Eur. Radiol. 31 (8) (2021) 6096–6104.

[27] L. Li, L. Qin, Z. Xu, et al., Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT[J], Radiology (2020).

[28] Mangal A, Kalia S, Rajgopal H, et al. CovidAID: CO‘VID-19 detection using chest Xray[J]. arXiv preprint arXiv:2004.09803, 2020.

[29] X. Yu, S. Lu, L. Guo, S.-H. Wang, Y.-D. Zhang, ResGNet-C: A graph convolutional neural network for detection of COVID-19[J], Neurocomputing 452 (2021) 592–605.

[30] Akhilesh, Gadde Lohith Sai, et al. "Covid-19 Detection Using CNN Model with CT Scan Images." 2023 International Conference on Computer Communication and Informatics (ICCCI). IEEE, 2023.

[31] Reema, Gunti, et al. "COVID-19 EDA analysis and prediction using SIR and SEIR models." International Journal of Healthcare Management (2022): 1-16.

[32] Srinivas, Pvvs, et al. "Detection of COVID disease from CT scan images using CNN model." 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS). IEEE, 2022.

[33] J. Zhang, Y. Chu, N. Zhao, Supervised framework for COVID-19 classification and lesion localization from chest CT[J], Ethiopian J. Health Dev. 34 (4) (2020).