iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.78916

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Hybrid Model for Early Diagnosis of Interstitial Lung Disease

Ms. N Lakshmi Deepthi, M Madhavi, S Naga Praneetha, M Pravalika
Department of CSE (Data Science) Institute of Aeronautical Engineering Hyderabad, India

Abstract: ***Interstitial Lung Disease (ILD) is a collection of progressive pulmonary conditions that impact on the pulmonary tissue structure and progressively diminish respiratory function. Early identification of ILD is still a difficult task since radiological differences are delicate and may be confused with other respiratory diseases. In the recent past, artificial intelligence has been developed to analyze medical images automatically and this has offered a new chance of enhancing diagnostic accuracy. The proposed study will develop a hybrid framework of deep learning based on Vision Transformer (ViT) and Convolutional Neural Network (CNN) structures to performearly detecting of ILD using chest CTscansand X-rays. The CNN element is concerned with the extractionoffine-scale localspacefeatures,including the texture anomalies and fibrotic structures, the Vision Transformer is concerned with the global contextual relationship between lung regions and other regions through the self-attention mechanisms. The representations extracted are combined to come up with an integrated prediction model that can differentiate between lungs with ILD and normal ones. A web-based clinicalsupport system iscreatedto facilitatethe real-time predictionbygivingmedicalpractitionersanopportunity to post-imaging data and receive automated diagnostic information. As shown in the experiments, the hybrid architecture suggested is better at classification than the single-model solutions, especially on the detection of early-stage abnormalities.***

Keywords: InterstitialLungDisease,VisionTransformer, CNN, Medical Imaging, Deep Learning, Computer-Aided Diagnosis.

I. INTRODUCTION

InterstitialLungDisease(ILD)isawiderangeofpulmonary disorders, which aremarkedby inflammationandfibrosis of interstitial lung tissues. Such structural changesdecrease the ability to exchange oxygen and this can eventually cause respiratory failure unless the condition is spotted early. In spite of the advancement in the radiology field, the area is challenging because the early identification is not possible sincethesymptomsare similartootherrespiratorydisorders likethecontinuouscough,fatigue,andbreathlessness.Useof medical imaging methods such as High-Resolution ComputedTomography(HRCT) andchestX- raysiscrucial in the diagnosis of ILD. Manual interpretation is however a very skilled process and subject to inter-observervariability. Minimal visual configurations that can occur in early stages of the disease are usually missed leading to late treatment. Deep learning has become an influential instrument of automatic analysis of images. Convolutional Neural Networks (CNNs) have been found to perform well in the identification of local visual features of medical images. Nevertheless, CNNs are mostly local to receptive fields, and they might not be able to provide long-range dependencies over the whole lung structure.

VisionTransformers(ViTs)haverecentlyproventobemore effective in capturing global correlations in images with the help of attention. In contrast to CNNs, transformers process image patches at the same time, which helps them to get a clearer understanding of the image context. These works inspiredthisworktosuggestahybridViT-CNNarchitecture, whichmixesthelocalfeatureextractionand globalcontextual learning methods to boost the accuracy of early ILD detection. The system proposed is meant to assist clinicians by offering an intelligent decision support tool that can help to provide fast and reliable diagnosis. This inconsistency results in the delayed diagnosis, misinterpretation, and manual evaluation inconsistencies. Along with the rapid advancement of artificial intelligence, deep learning has become a promising solution to the analysis of intricate medical data and the removal of the restrictions of the traditional diagnostic tools. CNNs have shown impressive performance in identifying visual patterns in healthcare images, and thus, they are appropriate to identify the ILD- related abnormalities. Nevertheless, the imaging data might notbeabletogiveacompletedepictionoftheseverityofthe disease; the presence of clinical data that includes the outcomes of the spirometry, patient demographics, oxygen saturationlevels,andmedicalhistoryhasapivotalroleinthe diagnosis. Conventional AI networks which utilize single- mode data do not tend to capture such complex connections.

©IJRASET: All Rights are Reserved

II. LITERATURE REVIEW

Initial work in the diagnosis of ILD was based on manually created image features and classical machine learning procedures. As the idea of deep learning progressed, CNN- related methods took over the deep learning-based image classification problem.

Anumberofstudiesprovedtheusefulnessof CNNstodetect lungpatterns(ground-glass,reticulation,andhoneycombing) inCTs.Themodelscontributedgreatlytoclassifyingfeatures better than the classical methods of texture-analysis To overcometheselimitationshybriddeeplearningarchitectures were subsequently proposed and involve the combination of several neural network models. Earlier studies combined CNNswithrecurrentneuralnetworks(RNNs) to achievethe input of serial clinical data and imaging data. Even though these methods enhanced the predictive performance, they neededorganizedtemporal datasetsandmorecomplex method makes it possible to detect better, especially at early stages of ILD where the abnormalities are mild and spread across space.

III. DATA COLLECTION AND PREPROCESSING

Thequalityofthedataandtheprocessofpreprocessingisof critical importance to the performance of deep learning models.Thedataemployedinthepresentstudyisasetofthe chest CT scans and X-ray images accessed through the publicly available reposit ries of medical imaging and curated data The quality, diversity, and consistency of the dataset on which the deep learning models are trained and evaluated heavily contribute to the performance of the models. In this research, medical imaging information was gatheredinrepositoriesontheInternet,whichcontainedchest CT scan and chest X-ray images of different patterns of Interstitial Lung Disease (ILD). The data sets consist of normallungconditionsanddiseasedcasesthathave Systems day,moretransformer-basedmodelsareusedin abnormalitiessuchasfibrosis,ground-glassopacities,andcomputervision activities.VisionTransformers make use of self-attention to learn the image dependencies of the world, which allows learning representations better. Medical imaging research has indicated that transformer models achievehighersuccessincomplexspatialrelationscompared tooldCNNs.Regardlessofthesedevelopments,therearefew studies that investigate the possibility of using CNNs and Vision Transformers together with the specific purpose of diagnosing ILD. This is the reason why the current piece of work can be described as a hybrid architecture that depends on CNN local feature sensitivity and ViT global contextual awareness to enhance pulmonary disease detection. Anthimopoulos et al. created a deep Convolutional Neural Network (CNN) to categorize High-Resolution Computed Tomography (HRCT) images of lungs into the various patterns of Interstitial LungDisease (ILD).Thearticle based their researchon abigdatasetofover16,000 imagepatches, and it allowed the model to identify subtle radiographic changes that are connected to various lung disorders. The system brought a huge workload off the shoulders of radiologists with automated analysis of pixel-level fibrosis andaccuracyofpatternrecognition.Nonetheless,ithasbeen demonstratedthatthemodelwasnotverygoodatidentifying the cases of ILD at the initial stages, when the fibrosis patterns are few or visually insignificant. This weakness emphasized the fact that models that are based on localized features of images can be less effective in capturing larger structural associations that occur in the initial stages of diseasedevelopment.VisionTransformer(ViT)modelshave also becomea promising alternative choicebecause they are abletocaptureglobalcontextualinformationintermsofself- attentionmechanisms.Transformersextractttherelationships between remote areas of the image in contrast to the traditional CNNs, which are more based on the locality of feature extraction, and this enables the distributed lung abnormalitiestobebetterunderstood.Inspiredbytheresults, the designed study will use a hybrid method of learning that integrates CNN and Vision Transformer architectures. The CNN element is effective at capturing fine-grained local texture patterns like reticular structure and variations in opacitiesandtheViTonecapturesglobalrelationshipsacross thelungimageasawhole.Thiscomplementinglearning reticular structures. These data sets include labeled medical imagesthathave been taken under varying imaging settings, imagingresolutions,andclinicalsettings,whichbringssome variability that should be overcome prior to model training. Inordertoachieveuniformity,alargepreprocessingpipeline was used. Firstly, all images were made to be resized to simpler sizes so that they could fit in the deep learning architecture and so that their computation might be simplified.Normalizationofpixelswasthendonetouniform intensity values to a uniform range and this enhanced convergenceofthemodelduringtraining.Filteringwasused as noise reduction technique to eliminate imaging artifacts andimprovesignificantstructuraldetails.Data-augmentation techniques,suchasrotation,horizontalflipping,scaling,and small manipulations, were used in order to make datasets more diverse and enhance model generalization and reduce overfitting. Moreover, the improvement of the lung region methods were implemented to highlight the clinically significant object and to reduce the irrelevant background data to enable the model to underline the pathological features.The datasetwasthenpreprocessed ,followedby the balanced coping of the data into training, validation and testingsubsets,wheretherewerenodata leakagesacrossthe sets.

This progressive output facilitates reduced variation in resolution, direction, and contrast, which assists the hybrid CNN -Vision Transformer model to learn more robust features and act more consistently in evaluation. score andArea Under the Receiver Operating Characteristic Curve(AUC-ROC)areusedtoevaluatemodelperformance,

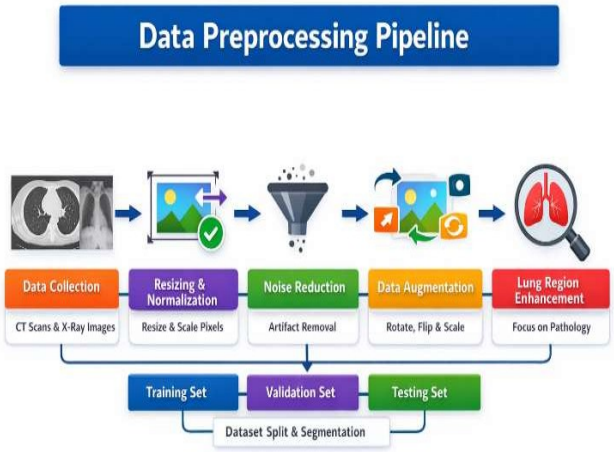


Fig.1: Dataprocessingpipeline

IV. SYSTEM DESIGN AND IMPLEMENTATION

Theproposedsystemisahybriddeeplearningmodel,which combines Convolutional Neural Network (CNN) andVision Transformer (ViT) models to allow the detection of InterstitialLungDiseasewithmedicalimagestobeaccurate. Thegeneralworkflowconsists of image acquisitioninwhich chest CT scans and the X-ray images are taken and made ready to be analyzed. The images are preprocessed to normalize the quality of inputs and then forwarded to the feature extraction phase. The CNN segment centers on the acquisition of local information such as edges, textures and patterns associated with fibrosis through the utilization of layers that sequentially identify finer abnormality in the lungs. Simultaneously,theVisionTransformerprocessesthe entire image splitting it into tiny blocks and with the help of attentionhowvariouspartsofthelungsareconnectedtoone anotherThis ability allows the model to identify distributed diseasepatternsthatarenotnecessarilyeasytoidentifyusing onlyconventional convolution operations BothCNN andViT branchesproducefeaturesthatarethencombinedonafeature fusion approach in which extracted representations are combined and fed through fully connected layers to learn built-infeatureinteractionsandthefinalclassificationoutput is produced. Python is the language in which the system is implemented along with deep learning frameworks like TensorFlow or PyTorch and OpenCV does the image processing operations.

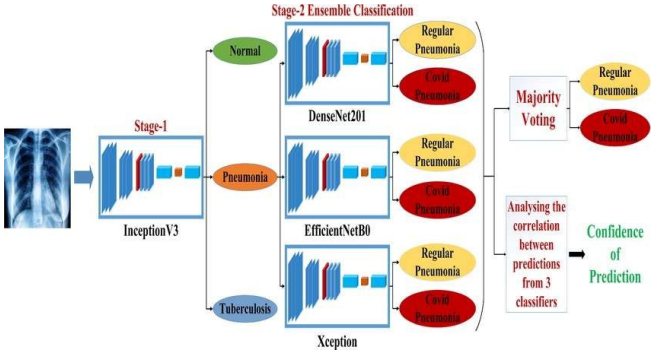


Fig.2:ConvolutionalNeuralNetwork(CNN)architecture

A python-based backend service developed with Flask or FastAPI manages model inferences and communication via APIs and a frontend based on React.js offers a convenient web-based interface where clinicians can submit medical pictures and receive real-time predictions. The proposed hybrid framework will be based on feature representations trained on both the Convolutional Neural Network (CNN) and Vision Transformer (ViT) branches to produce a single diagnostic prediction. The extracted features of the two architecturesarecombinedtogetherinafusionlayerandsent to the fully connected layers to be trained to learn the combinedfeatures.

Thelastoutputlayergeneratestheresults of classification that show whether Interstitial Lung Disease (ILD) is present or absent with optional prediction of the severitylevelsofthediseases.Intraining,AdamWoptimizer, dropout regularization, batch normalization and early stopping are used as optimization techniques to enhance convergencestabilityandavoidoverfitting.Standardresults, suchasaccuracy, precision,recall,sensitivity,specificity,F1- and exhaustiveness of diagnostic Python languages are used to implement the system that applies deep learning frameworks (PyTorch or TensorFlow) and other libraries (OpenCV and pydicom) to process medical images and DICOM files. Other preprocessing and evaluation tools are assisted with the help of the scikit-learn library. Upon training, the model is deployed in a backend application programmed in FastAPI or Flask but this exposes RESTful interfaces through which clinicians can send CT or X-ray images that will be analyzed automatically. The backend is the input validation, preprocessing, model inference and prediction result storage, which stores prediction results in databaseslikePostgreSQLorMongoDB,withlargeimaging files and model artifacts being stored with object storage systems like Amazon S3.

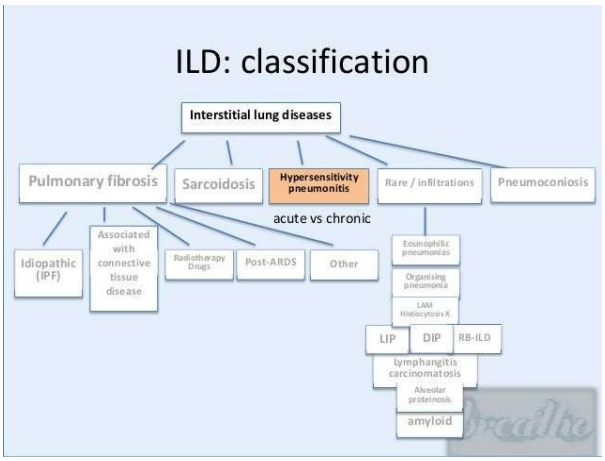


Fig.3:Classification

An online interface created with React or Vue helps the clinician to log in safely and upload patient imaging information and see the prediction probability and interpretability results in the form of Grad-CAM heatmaps showing areas that can make the model make decisions.The whole system is containerized with Docker to maintain a consistentdeploymenttoallenvironmentsandisdeployedto the cloud with support of eitherAWS or Google Cloud, and with a provision of the enabled use of a GPU to speed up prediction. Such security measures as HTTPS encryption, role-based access control, audit logging, and anonymization of patient identifiers are included to ensure privacy of healthcaredata.Performancedriftwithtimecanbeobserved with the help of monitoring, logging, and version control mechanisms whereas the security of consistent system updates can be achieved with the support of automated continuous integration pipelines.Altogether, the worked-out framework is a full-fledged end-to-end AI pipeline that can provideeffective,scalable,and explainableILDdiagnosisto be applied in real-world clinical practice.

V. RESULTS AND DISCUSSION

To comprehensively evaluate the diagnostic effectiveness, the proposed hybrid Vision transformer-Convolutional Neural Network (ViT-CNN) model was tested by means of the standardclassificationperformancemetrics,suchas accuracy, precision, recall, F1-score, and Area Under the ReceiverOperatingCharacteristicCurve(ROC-AUC).Ithas been experimentally revealed that CNN-only models demonstrated capability in identifying salient lung abnormalities by learning localized texture and structural characteristics, but demonstrated weaknesses in global contextual relationships among the lung areas. On the other hand,ViT-onlymodelswerecapableofmodelinglong-range spatial dependencies and general structural patterns though this model needed more specific featureguidanceto identify abnormalities accurately. The proposed hybrid model was effective at tapping complementary strengths by combining both architectures, and this took the form of better overall classification performance as compared to single models.

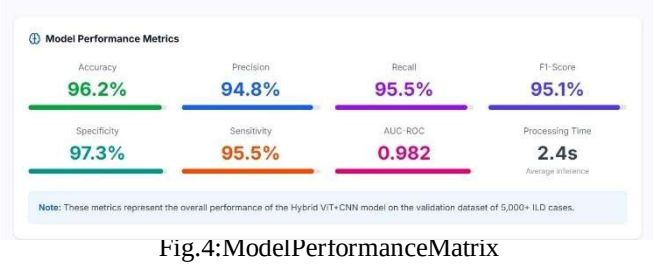


Fig.4:ModelPerformanceMatrix

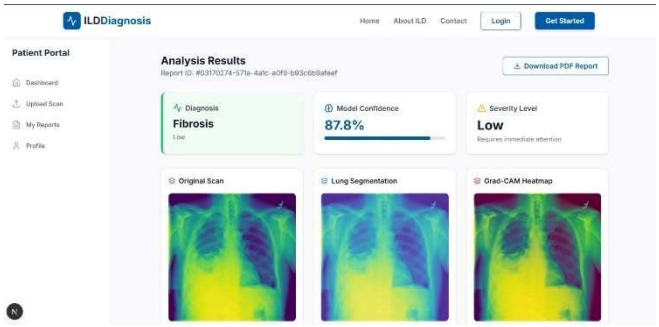


Fig.5:AnalysisResults

It is interesting to note that the hybrid strategy demonstrated greater sensitivity in detecting the cases of ILD at an early stage, and thus, false-negativeexpectationswereminimized, which is a paramount requirement in medical diagnosis as treatment may be postponed in cases of missed diagnosis. Moreover,attention-mechanism-basedvisualizationmethods andGrad-CAMheatmapsshowedthatthemodelconsistently concentrated on clinically significant lung areas of fibrosis and abnormal tissue patterns, which allowed to improve the interpretability and boost clinical trust in automated inferences. It was seen that the training process showed consistent convergence with less validation loss and better generalization on unseen test samples, which indicated that theproposedarchitectureisrobust.Allinall,theresultsofthe experiments prove that the local feature extraction with a global contextual learning isa morereliable and explainable framework used to detect early ILD and can be used as the computer-aided diagnostic tool in the real-life healthcare setting.



Fig.6:Confidencelevel

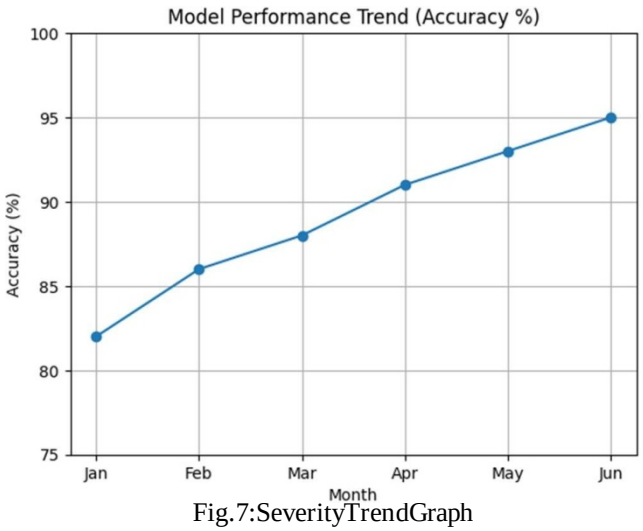


Fig.7:SeverityTrendGraph

Besides quantitative analysis, qualitative analysis was done through interpretability methods, including Grad-CAM and attention visualization, to gain a better insight into how the modelmakesdecisions.
TheproducedheatmapsofbothCNN and Vision Transformer branches showed that the model always attended to areasof clinical interest in the lung, such as those exhibiting fibrosis, reticulation, and the patterns of opacities, but not to non-informative body parts such as the ribs and tissues. This action enhances the transparency and gives the clinicians more confidence in the automated predictions. Testing in a simulated deployment environment found that the system delivered effective inference performancewithpredictionsresultstakingafewsecondsto run on systems with GPUs and the system being able to acceptably respond in seconds on systems with computers. Thewebapplicationdevelopedenablesclinicianstopost CT or X-ray images, see probability-based predictions and analyze visual explanation maps, thus, supporting informed clinical decision-making.

VI. CONCLUSION AND FUTUREWORK

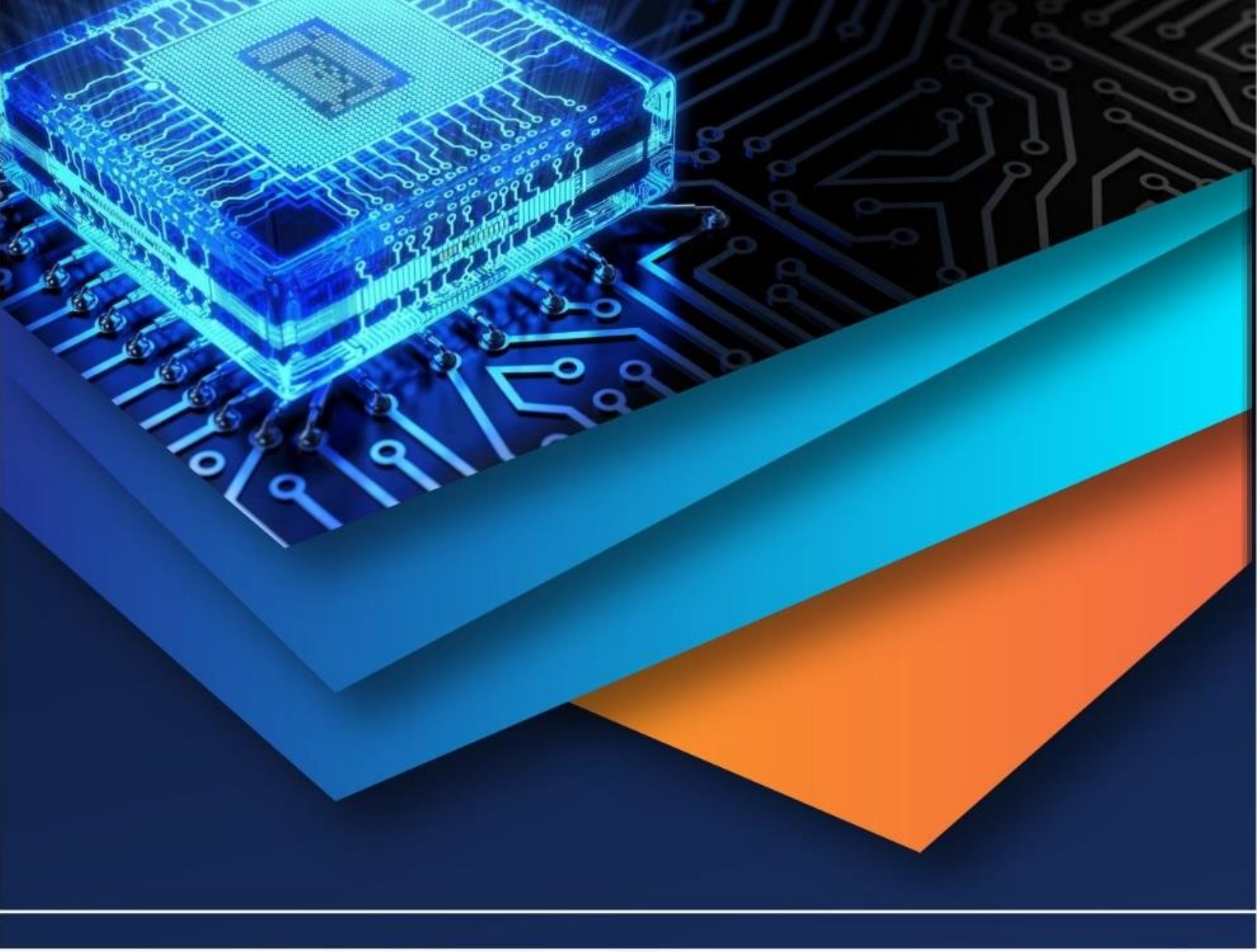
Through the integration of CNN-based local feature extraction with the global contextual learning potential of transformer models, the suggested methodology shows a better diagnostic performance with respect to traditional single-model methods. The hybrid design allows making betterdetectionoffaintlunganomaliesespeciallyintheearly stages of diseases when the visual patterns can be very hard to discern during the manual assessment. The designed system is also an intelligent clinical decision-supportsystem which can deliver prompt, predictable, and interpretable predictions based on a web-based interface, thus helping healthcareprofessionalsto enhancediagnosticeffectiveness and decrease analysis time. Experimental assessment proves the efficiency, strength, and functionality of transformer- based hybrid learning in terms of medical image analysis contexts. Regardless of the positive outcomes, there are a numberoffutureimprovementopportunities.Futureresearch willinvolvetheextensionofthemodeltoscanentirevolumes of full 3D CT scan to obtain more spatial information and more accurate detection of complex ILD patterns. Improved attention-basedfusionmethodscanbeinvestigatedtohavean evenbetterfeatureintegrationbetween CNNandtransformer elements. Increasing the data by using multi-institutional cooperationwillassistinenhancingthemodelgeneralization toawiderrangeofpatientsandimagingconditions.Also,the framework can be expanded to accommodate multi-disease classification to identify other pulmonary diseases like pneumonia,COPDandlungcancer.Additionalstudiescanbe conducted as well, such as directly validating in real-time clinical, lightweight model optimisation as an edge deployable model, and integration with electronic health record to implement continuous and scalable AI-assisted healthcare solutions.

REFERENCES

[1] J.Li,J.Chen,Y.Tang,C.Wang,B.A.LandmanandS.K. Zhou, Transforming medicalimagingusing transformers? Comparison Review of essential Properties, up-to-date Advances, and Future Proximalities, Medical Image Analysis, vol. 85, p. 102762, 2023.
[2] Y.Zhang,J.Wang,J.M.GorrizandS.Wang,"Deep Learning and Vision Transformer for Medical ImageAnalysis,"JournalofImaging,vol.9,no.7,p.147, 2023.
[3] A.Halder,S.Gharami,P.Sadhu,P.K.Singh,M.Woźniak and M. F. Ijaz, "Implementing Vision Transformer for Classifying 2D Biomedical Images," Scientific Reports, vol. 14, 2024.

[4] Sarmadi, Z. S. Razavi, A. J. van Wijnen et al., "Comparative Analysis of Vision Transformers and Convolutional Neural Networks in Osteoporosis DetectionfromX-rayImages,"ScientificReports,vol.14, Art. no. 18007, 2024.
[5] J. Zhang, F. Li, X. Zhang, H. Wang and X. Hei, "AutomaticMedicalImageSegmentationwithVision Transformer,"AppliedSciences,vol.14,no.7,p.2741,2024**.**
[6] S.Raminedi,S.ShrideviandD.Won, "Multi-Modal Transformer Architecture for Medical Image Analysis and Automated Report Generation," Scientific Reports, vol. 14, Art. no. 19281, 2024..
[7] Halder, S. Gharami, P. Sadhu, P. K. Singh, M. Woźniak and M. F. Ijaz,"Implementing Vision Transformer for Classifying 2D Biomedical Images," Scientific Reports, vol. 14, Art. no. 12567, May 2024.
[8] Hybrid Vision Transformer Architectures with CNN Blocks for Multi-Label Chest Disease Classification," Power System Technology Journal, vol. 49, no. 1, Apr. 2025.
[9] J.Qezelbash-ChamakandK.Hicklin, "A Hybrid Learnable Fusion of ConvNeXt and Swin Transformer for Optimized Image Classification," IoT, vol. 6, no. 2, p. 30, May 2025.
[10] Safdar and M. Saadeldin, "CoMViT:AnEfficientVisionTransformerBackbonefor Supervised Classification in Medical Imaging," arXiv preprint arXiv:2510.27442, 2025.
[11] J.W.Kim,A.U.KhanandI.Banerjee, "Systematic Review of Hybrid Vision Transformer Architectures for Radiological Image Analysis," Journal of Imaging Informatics in Medicine, vol. 38, pp. 3248– 3262, 2025..

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)