Multi-Disease Diagnostic Framework Using VGG19-Based Deep Visual Feature Extraction from Medical Images

Manojkumar G [1], *David Neels Ponkumar* D[2], *Nithya*.S[1], *Selvakani* R[2], *Vidhyapathy*.M[1] and *Balakrishnan*.N[1]

[1]*Department of Information and Technology, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College* Chennai-600062, India
[2]*Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology,* Chennai-600062 India

Abstract. Effective clinical decision-making depends on the early and accurate diagnosis of thoracic illnesses. In order to analyze chest X-ray and CT pictures, this work proposes a deep learning-based multi-disease diagnosis system that uses transfer learning with the VGG19 convolutional neural network. The suggested model is trained on publically accessible datasets that have undergone extensive preprocessing, such as data augmentation, scaling, and standardization to improve robustness. According to experimental findings, training accuracy steadily increases to about 70%, while validation accuracy stays steady at about 50%, suggesting that dataset complexity and inter-class similarity limit generalization. While the validation loss exhibits minor variations, indicating the existence of moderate overfitting, the training loss generally displays a decreasing trend. In spite of this, the framework successfully acquires discriminative characteristics pertinent to the categorization of lung diseases. The suggested method supports dependable and data-driven healthcare decision-making by showcasing the potential of transfer learning for medical imaging applications and offering a scalable basis for AI-assisted diagnostic systems.

1 Introduction

Healthcare diagnostics has seen a dramatic change as a consequence of the usage of Deep Learning (DL) and Artificial Intelligence (AI) to medical imaging. Conventional diagnostic methods that only use radiologists are frequently limited by inter-observer variability, workload, and subjectivity. More precise, reliable, and automated disease diagnosis across a variety of imaging sense modality, including X-ray, CT, MRI, and fundus imaging, has been made possible by the development of data-driven AI algorithms [1,2]. Study in this area has been further accelerated by the exponential rise of annotated medical datasets and the accessibility of high-performance computing systems. Due to its ability to learn hierarchical representations of spatial and textural elements from picture data, deep learning styles in particular, Convolutional Neural Networks (CNN) have emerged as the mainstay of automated medical image interpretation [3,4]. Even with slight domain specific data, pre-trained models like VGG19 and ResNet can be efficiently modified for medical imaging applications through the use of transfer learning [5]. This method is appropriate for real-world clinical deployment since it maintains excellent accuracy while cutting down on training time and computing expense. Multi disease analytical frameworks that can find several illnesses from several imaging modalities.

Selvakani.r@gmail.com

A deep learning-based multi-disease diagnosis model called Health Scan AI. It uses feature fusion and transfer learning to improve performance in multi-class medical picture classification. These advancements show how comprehensive and integrated diagnostic models have replaced conventional single-disease detection techniques. One of the best CNN models for medical image processing is still the VGG19 architecture, which was pretrained on the ImageNet dataset. It can effectively generalise across a variety of domains thanks to its deep hierarchical structure and uniform layer design. In order to increase transparency and clinician confidence in automated predictions, Behera et al. [6] used a VGG19-based system for pneumonia identification with Explainable AI (XAI) integration. In medical AI systems, explainability techniques are especially crucial since doctors need both the forecast result and an understandable justification for their diagnoses.

Explainable AI has been used in a variety of medical settings, not just the identification of pneumonia. The viability of multimodal techniques in healthcare diagnostics was demonstrated in [7], who investigated tongue imaging for non-invasive disease diagnosis utilising explainable modelling and deep learning. The trend towards interpretable and multimodal diagnostic systems in contemporary healthcare was also highlighted by [8], who used multimodal imaging and explainable AI approaches for the detection of skin diseases. The accuracy and resilience of AI models in medical imaging have been substantially improved by attention-based deep learning techniques. It created a cross-modal attention-based model that combines abdominal X-ray images with symptom questionnaires to diagnose defecation, suggested a framework that incorporates attention mechanisms and transfer learning to improve diagnostic precision in X-ray imaging. By using multimodal and attention-guided techniques, models are able to replicate the comprehensive thinking of medical practitioners by capturing intricate correlations between clinical symptoms and picture elements [9].

Improvements in medical imaging hardware have made it easier to obtain higher-quality data, which is crucial for deep learning applications, in tandem with algorithmic breakthroughs. Using perovskite-based detector materials, improved the sensitivity of X-ray detection and used common CMOS sensors to demonstrate high-resolution phase-contrast X-ray imaging [10]. By improving image clarity, these advancements support AI-driven models and boost diagnostic performance. To diagnose many diseases from medical photos, a multi-stage deep learning system is suggested. To enhance feature generalisation and lessen overfitting, the system uses transfer learning, data augmentation (rotation, flipping, shifting, and zooming), and incremental fine-tuning of the VGG19 model. System of measurement with correctness, precision, recall, and F1-score to assess the model once it has been trained and validated on benchmark medical image datasets, such as the Kaggle Chest X-ray (Pneumonia) dataset. In clinical contexts, the trained model tin be used by way of a choice provision tool to provide AI-assisted picture interpretation for conditions like lung cancer, pneumonia, and tuberculosis.

The suggested architecture is in line with new developments in AI systems that use less resources. The feasibility of deep learning-based breast cancer malignancy prediction on embedded devices, highlighting the potential for portable diagnostic models in healthcare settings with limited resources. Therefore, the suggested solution seeks to close the gap between clinical usefulness in the actual world and high-performance AI models. The foundation for next-generation intelligent diagnostic systems is established by the combination of multimodal imaging, explainable AI, and deep learning [11]. To enable reliable multi-disease diagnosis, the suggested model makes use of VGG19 fine-tuning, transfer learning, and data augmentation.

2 Literature Review

Because deep learning makes it possible to accurately and automatically interpret complex radiographs, it has become essential in X-ray-based multi-disease diagnosis. In order to improve feature embedding and detect co-existing pathologies, Xing et al. [1] presented a multi-label proxy metric learning framework for multimorbidity recognition and retrieval from X-ray datasets. By using attention processes and robust transfer learning on medical X-rays, Das et al. [6] were able to improve model focus on discriminative illness regions and achieve higher accuracy and resistance against noise. Sangnark et al. [7] improved interpretability and cross domain feature interpretation by integrating cross modal attention between abdominal X-ray pictures and clinical questionnaires to diagnose dyssynergy defecation.

Table.1 Overall study process of existing system

Ref.	Authors & Year	Imaging Modality / Dataset	AI / Deep Learning Technique	Drawback
[1]	Y. Xing et al., 2023	General medical X-rays (multi-disease dataset)	Multi-label Proxy Metric Learning	Limited interpretability and lack of real-world clinical validation
[4]	S. R. Chokhandre et al., 2024	Clinical breast X-ray (mammography)	CNN model optimized for embedded devices	Reduced model complexity led to minor drop in diagnostic sensitivity
[6]	I. Das et al., 2025	Chest X-ray (NIH, CheXpert)	Attention Mechanisms + Robust Transfer Learning	High computational cost and dependence on pretrained backbones
[7]	S. Sangnark et al., 2024	Abdominal X-ray + clinical questionnaire	Cross-Modal Attention Deep Network	Requires multimodal input; limited to small clinical cohort
[8]	V.-N. Pham et al., 2023	Fundus & X-ray images (multi-modal)	Consideration Aided Generative Learning (GAN)	Generated images may not fully preserve clinical fidelity
[11]	S. Mandot et al., 2024	X-ray Fluorescence Emission Tomography (XFET)	AI-assisted reconstruction and imaging-spectrometer array	Complex hardware setup and high system cost
[12]	M. Pérez et al., 2025	Phase-contrast X-ray using CMOS sensors	AI-enhanced image acquisition and processing	Hardware-dependent method; limited clinical dataset for validation

Chokhandre et al. [4] proved that AI deployment is feasible in low-power healthcare devices by creating a resource-efficient embedded system for breast X-ray malignancy diagnosis. For multi-scale, multi-modal fundus and X-ray image translation, Pham et al. [8] presented an attention-aided generative learning model that supports cross-modality learning and data augmentation. For better illness localization, Mandot et al. [11] constructed a high understanding bench-top X ray fluorescence emission tomography (XFET) device with AI-assisted reconstruction. Pérez et al. [12] improved low-dose imaging and produced high-quality data for deep learning analysis by utilizing commercial CMOS image sensors for phase-contrast X-ray capture. The integration of attention, generative learning, and embedded AI models improves diagnosis accuracy, interpretability, and deployment efficiency in X-ray-based multi-disease healthcare systems, as this research together show [Table.1].

Research Gap : Current X-ray AI models perform well on small datasets but are unable to generalize to other illnesses and medical facilities. Doctors find it difficult to trust or understand AI-based diagnosis because most systems are not easily explained. Deep models require lightweight, effective designs because they are frequently too hefty for real-time or low-power clinical devices.

3 Proposed Methodology

In Before In order to accurately detect pneumonia from chest X-ray pictures, the suggested method presents a multi-stage deep learning framework that makes use of gradual fine-tuning of the VGG19 architecture and transfer learning [1]. The subsequent are the main basics of the arrangement workflow [Fig.1].

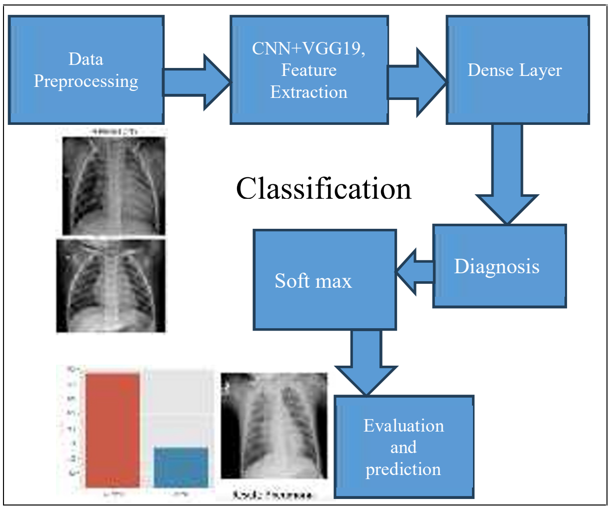

Fig.1 Overall proposed system architecture

Obtaining and Preparing Data the Kaggle Chest X-ray (Pneumonia) dataset, which includes annotated images of both normal and pneumonia patients, is used by the system. Every image is normalized, scaled to 128 × 128 pixels, and RGB-converted. Rotation, flipping, shifting, and zooming are examples of data augmentation techniques that increase dataset diversity and lessen overfitting [2]. Architecture Model the VGG19 convolutional neural network, pretrained on ImageNet and featuring fully connected layers tailored for binary classification, serves by way of the foundational feature separator. In instruction to precisely detect pneumonia from chest X-ray pictures, the suggested method presents a multi-stage deep learning framework that makes use of gradual fine-tuning of the VGG19 architecture and transfer learning [3].

The succeeding are the chief elements of the arrangement workflow: Obtaining and Preparing Data the Kaggle Chest X-ray (Pneumonia) dataset, which includes annotated images of both normal and pneumonia patients, is used by the system. Every image is normalized, scaled to 128 × 128 pixels, and RGB-converted. Rotation, flipping, shifting, and zooming are examples of data augmentation techniques that increase dataset diversity and lessen overfitting. Architecture Model

The VGG19 convolutional neural network, pretrained on ImageNet and featuring fully connected layers tailored for binary classification, serves as the foundational feature extractor [4].

A dependable decision support system for doctors can be obtained by participating the final model, which is saved as vgg_unfrozen.h5, into a medical analytic web or mobile application for AI assisted chest X-ray processing.

3.1 Data Preprocessing

Step 1: Extracting Datasets: This step involves utilising the Python zip file module to extract the compressed dataset that contains medical X-ray images from a ZIP archive. The files that have been extracted are saved in a designated directory for later use. Through this procedure, all data is guaranteed to be organised in a consistent file hierarchy that is appropriate for automated processing. Although it doesn't need any mathematical calculations, this stage serves as the basis for repeatable data loading and management in deep learning pipelines [5].

Step 2: Listing Image Files: All image files in the extracted directory with the extensions.jpg,.jpeg, or.png are found using a recursive file search. The os. walk () function creates a list of acceptable image paths by iterating over subdirectories. By preventing errors in human data treatment, this stage guarantees thorough dataset inclusion. Despite being entirely computational, it preserves the integrity of the dataset by guaranteeing uniform input coverage for subsequent preprocessing and analysis phases.

Step 3: Converting images to greyscale and loading them: Each image is converted into a single-channel greyscale format by loading it into memory using OpenCV's cv2.imread() method and the IMREAD_GRAYSCALE parameter. Colour information is unnecessary because X-ray imagery mostly depicts intensity-based tissue densities. The image is represented mathematically as $I\ (x,\ y)\ \in$ [0,255], where the radiography attenuation coefficient is encoded by each pixel intensity. This conversion preserves diagnostically significant features while lowering computational complexity.

Step 4: Using Otsu's Thresholding to Extract the Skull Region First, a 5 × 5 kernel is used for Gaussian smoothing in order to decrease high-frequency noise and separate the skull region. The hazy picture

$$I_b(x,y) = I(x,y) * G(x,y;\sigma) \qquad (1)$$

is thereafter exposed to Otsu's binarization technique. By minimising the intra-class variance of pixel intensities, Otsu's approach determines the ideal global threshold t^*, stated where ω_i and σ_i represent variances and class probabilities, respectively. By differentiating high-intensity bone features from the background, the binary mask that is produced isolates the skull. A 7×7 kernel is then used to apply a morphological closing operation $A \cdot B\ =\ (A \oplus B)\ \ominus B$ in order to remove tiny holes and fine-tune mask boundaries. Element-wise multiplication is used to obtain the last portion of the skull [9].

$$I_{skull}(x,y) = I\,(x,y).\,M(x,y) \tag{2}$$

where the binary mask is denoted by $M\,(x,\,y)$.

Step 5: RGB conversion and image resizing: Bilinear interpolation is used to resize the isolated skull image to 224 × 224 pixels, guaranteeing compliance with the VGG19 input specification. The intensity channel is then duplicated across R, G, and B dimensions to transform the greyscale image to a three-channel RGB format,

$$I_{rgb}(x,y,c) = I_{gry}(x,y) \tag{3}$$

for c ∈ {R, G, B}. For transfer learning applications that use pretrained convolutional neural networks, this process standardises the input size and format.

Step 6: VGG19 Input Normalisation: Per-channel normalisation is performed to match the image distribution to the ImageNet dataset used for VGG19 pretraining. Initially,

Pixel intensities are converted to the [0, 1] range using the formula

$$I_{scaled} = {}^{I_{rgb}}/_{255} \tag{4}$$

The VGG19 normalisation constants are then used to accomplish channel-wise standardisation:

$$I_c'(x,y) = \frac{I_{scaled}(x,y,c)-\mu_c}{\sigma_c} \tag{5}$$

in which μ = [0.485,0.456,0.406] and σ = [0.229,0.224,0.225] are the ImageNet dataset's empirical mean and standard deviation, respectively. By guaranteeing uniform feature scaling, this enhances convergence stability while fine-tuning the model [6].

Step 7: Verification and Visualisation: The original greyscale X-ray, the extracted skull mask, the enlarged RGB picture, and the normalised image tensor are the four visualisation panels that are created to examine the intermediate outcomes. Visualisation facilitates qualitative validation of ROI accuracy and preprocessing correctness. The normalised tensor is linearly rescaled to [0,1] for display reasons for interpretability, guaranteeing that all intermediate operations result in legitimate diagnostic representations.

Table.2 List of Pre Processing

Step	Operation	Key Formula / Concept	Output
1	Unzip dataset	—	Image folder
2	Collect images	—	List of file paths
3	Read grayscale	[illegible]0,255]	Gray image
4	Skull ROI (Otsu + Morphology)	$I \in [0,255]$ $t^* = \arg\min_t \sigma_w^2(t)$	Skull mask
5	Resize + RGB	[illegible] = [[illegible]	(224×224×3)
6	Normalize (VGG19)	$I_{rgb} = [I, I, I]$ $I' = (I/255 - \mu)/\sigma$	VGG19 tensor
7	Visualization	Matplotlib	Display pipeline
8	Output	Final array shape	(224, 224, 3)

Step 8: Data and Output Preparedness: Dimensions of the final pre-processed image tensor are (224,224,3), and it is numerically standardised within roughly [−2, +2] [−2, +2]. For feature extraction or classification, this format works directly with VGG19 and other CNN-based architectures. The preprocessing pipeline thus provides a robust, standardized method for skull-region X-ray preparation, ensuring reproducibility, clinical relevance, and deep-learning compatibility in medical imaging studies [Table.2].

3.2 VGG19 + CNN classification algorithm

Step 1: Utilising pre-trained VGG19 for Feature Extraction The algorithm's initial step uses a pre-qualified VGG19 CNN to extract deep features. Small 3 × 3 × 3 convolutional filters are used in the 19-layer deep architecture VGG19, which preserves computational efficiency while capturing fine-grained spatial features. To maintain continuity with the network's initial training domain, the input medical X-ray or skull images are shrunk to 224 × 224 × 3 pixels and normalised using ImageNet mean and standard deviation values. After low-level edge and surface descriptors in earlier layers to elevated semantics abstractions of anatomical areas in later layers, each convolutional block in VGG19 carries out hierarchical representation learning. Critical activation properties are preserved as the spatial dimensions are gradually down-sampled by max-pooling processes. VGG19's convolutional filters already encode powerful general-purpose characteristics that can be successfully used to the medical imaging field because it has been pre-trained on a big dataset of natural images [14]. The model guarantees steady feature reuse and avoids overfitting by freezing these layers during fine-tuning. The final pooling layer's deep feature maps (7 × 7 × 512) yield a rich 25 088-dimensional descriptor vector for every image, which serves as the input for the dense-layer classifier that follows [7].

Step 2: Dense-Layer Mapping and Feature Flattening: To make learning by fully connected (dense) layers easier, the multidimensional convolutional feature maps generated by VGG19 are flattened into a one-dimensional feature vector in the second phase. This transformation allows matrix operations for high-level classification while preserving the spatially aggregated feature activations. To capture intricate inter-feature interactions, the flattened vector (of dimension 25 088) is fed into a series of thick layers that function as nonlinear projection spaces. Using 256 neurones with ReLU activation, the first dense layer reduces dimensionality and adds nonlinearity to the learnt representation. To lessen overfitting and enhance generalisation, a dropout layer through a amount of 0.5 is used to arbitrarily neutralize neurones through working out. The discriminative representation is further improved by the second dense layer of 128 neurones, which highlights prominent characteristics associated with particular disease groups [8]. To ensure quick convergence and stability, each dense layer uses backpropagation to optimise its weights using the Adam optimiser and categorical cross-entropy loss. In order to properly prepare the data for final probabilistic classification in the output layer, this hierarchical mapping technique converts raw visual descriptors into a condensed, task-specific feature space that is more separable across disease categories [9].

Step 3: Using Softmax for Classification and Decision Making: Using a Softmax output layer, the last step carries out classification by transforming the high-level dense-layer activations into normalised probability distributions for every illness category. To ensure that probabilities add up to one, the Softmax function mathematically calculates the exponential of each neuron's output and divides it by the total of exponentials across all classes. The probability for class I for an input vector z is provided by

$$P_i = e^{z_i} / \sum_j e^{z_j} \quad (6)$$

For gradient-based optimisation, this formulation preserves differentiability while enabling the network to express confidence in its predictions. A six-dimensional probability vector that corresponds to the predetermined illness categories such as normal, pneumonia, tuberculosis, etc. is produced by the model [10]. The class chosen for the projected diagnosis is the one with the highest probability value. System of measurement such by way of accuracy, correctness, recollection, and F1-score are calculated during evaluation to amount the system's performance. Additionally, threshold-based decision rules are supported by the probabilistic output for clinical applications where sensitivity is more important than specificity. The network reaches multi-class discriminating capability by combining Softmax with the previously mentioned dense representations. This translates intricate medical image patterns into comprehensible diagnostic results appropriate for automated computer-aided illness screening systems.

3.3 Pseudo Code

Algorithm: VGG19-Based Medical Image Classification

1. Begin
2. Load pre-trained VGG19 network (weights = ImageNet)
3. For each image I in dataset D:
4. a. Preprocess I:
   - Convert to grayscale (if needed)
   - Apply ROI extraction (Skull Region)
   - Resize I to 224×224×3
   - Normalize pixel values using ImageNet mean and std
5. b. Feature Extraction:
   - Pass I through VGG19 convolutional layers
   - Obtain deep feature map F (size 7×7×512)
6. c. Flatten feature map F into 1-D vector V (length = 25,088)
7. d. Classification Phase:
   - Input V into dense layer 1 (256 units, ReLU)
   - Apply Dropout (0.5)
   - Input to dense layer 2 (128 u, ReLU)
   - Output to Softmax layer (6 u)
8. e. Predict class:
   C = argmax (Softmax(V))
9. End For
10. Display class label and probability
11. End

3.4 Evaluation and Prediction

The efficacy and clinical dependability of AI models used for multi-disease diagnosis from medical images are largely determined by evaluation and prediction. Large and varied datasets are used to assess the model's performance following the training phase, confirming its capacity to generalize across various patient groups and imaging scenarios. Testing on unseen samples and cross-validation assist guarantee that the system can function reliably in actual clinical settings without overfitting to training data. In the prediction stage, new medical images like CT, MRI, or X-ray scans are analyzed by the trained deep learning architecture, which is often founded on CNN, Vision Transformers (ViTs), or hybrid fusion networks [11].

In order to identify several illness categories at once, the model automatically extracts structural and spatial data from the images. Low- and high-level visual properties are captured by feature extraction layers, and the final classification is carried out by fully linked layers or attention methods. Testing the model's flexibility to vicissitudes in image quality, noise, and class imbalance is another way to improve reliability. To ensure that the forecasts are in line with medical rationale, interpretability and clinical validation by subject-matter experts are equally crucial [12]. All things considered, integrating sophisticated feature extraction with methodical assessment guarantees that AI-driven diagnostic systems produce precise, comprehensible, and effective multi-disease prognosis, facilitating quicker and more accurate clinical decision-making.

4 Result and Discussion

4.1 Preprocessing output

Key processes in preparing medical pictures for deep learning-based disease diagnosis are demonstrated by the preprocessing result. The segmented region of interest known as the "Skull Region (ROC)" is where the skull area has been separated utilising morphological techniques and thresholding to separate significant bone features while eliminating background noise. Only anatomical traits that are pertinent to medicine are preserved thanks to this segmentation [Fig.2]. shows the chest X-ray after normalisation, where the image has been scaled, intensity-adjusted, and standardised to satisfy the input specifications of the VGG19 model are enhanced by the combination of these preprocessing techniques region extraction and normalisation [Fig. 3].

https://doi.org/10.1051/itmconf/20268203009

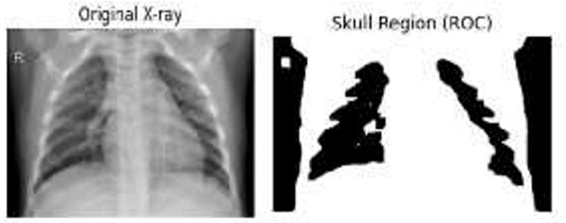

Fig.2 Normal image to convert ROC image

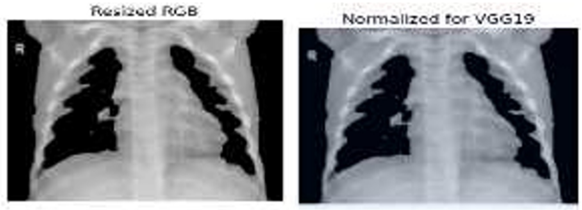

Fig.3 Input ready for normalized VGG19

4.2 VGG19 + CNN classification output

The VGG19 architecture used to classify medical images into six illness categories is shown in the result table. It begins with an input layer that can switch 224 x 224 x 3-layer images. Each of the five convolutional blocks that make up the model has several Conv2D layers, which are followed by MaxPooling2D layers. From 64 filters in the initial block to 512 in the deeper layers, the convolutional layers gradually increase the number of filters, enabling the network to extract low- to high-level textural and spatial information from medical pictures. MaxPooling decreases spatial dimensions, helping the model focus on key regions and minimising computational effort. The Flatten deposit transforms the 3-Dimnesional feature maps into a 1-Dimensioal vector with 25,088 pixels following the convolutional feature extraction process. Fully connected Dense layers (256 and 128 neurones) that discover intricate connections between the features that have been retrieved come next [Table 4]. During training, half of the neurones are randomly deactivated using a Dropout layer (0.5) to avoid overfitting. Lastly, the probability distribution across six disease classes is predicted by the output layer, which has six neurones and SoftMax activation.

Table.3 VGG19 result parameter

Layer (Type)	Output Shape	Parameters
Input-Layer	(None, 224, 224, 3)	0
Conv-2D (block1_conv1)	(None, 224, 224, 64)	1,792
Conv-2D (block1_conv2)	(None, 224, 224, 64)	36,928
MaxPooling-2D (block1_pool)	(None, 112, 112, 64)	0
Conv-2D (block2_conv1)	(None, 112, 112, 128)	73,856
Conv-2D (block2_conv2)	(None, 112, 112, 128)	147,584
MaxPooling-2D (block2_pool)	(None, 56, 56, 128)	0
Conv-2D (block3_conv1)	(None, 56, 56, 256)	295,168
Conv-2D (block3_conv2)	(None, 56, 56, 256)	590,080
Conv-2D (block3_conv3)	(None, 56, 56, 256)	590,080
Conv-2D (block3_conv4)	(None, 56, 56, 256)	590,080
MaxPooling-2D (block3_pool)	(None, 28, 28, 256)	0
Conv-2D (block4_conv1)	(None, 28, 28, 512)	1,180,160
Conv-2D (block4_conv2)	(None, 28, 28, 512)	2,359,808
Conv-2D (block4_conv3)	(None, 28, 28, 512)	2,359,808
Conv-2D (block4_conv4)	(None, 28, 28, 512)	2,359,808
MaxPooling-2D (block4_pool)	(None, 14, 14, 512)	0
Conv-2D (block5_conv1)	(None, 14, 14, 512)	2,359,808
Conv-2D (block5_conv2)	(None, 14, 14, 512)	2,359,808
Conv-2D (block5_conv3)	(None, 14, 14, 512)	2,359,808
Conv-2D (block5_conv4)	(None, 14, 14, 512)	2,359,808
MaxPooling-2D (block5_pool)	(None, 7, 7, 512)	0
Flatten	(None, 25088)	0
Dense (256 units)	(None, 256)	6,422,784
Dropout (0.5)	(None, 256)	0
Dense (128 units)	(None, 128)	32,896
Dense (Output = 6 classes)	(None, 6)	774

With over 20 million parameters, this VGG19-based model effectively extracts deep hierarchical features from medical images to enable precise multi-disease detection.

4.3 Prediction and Evaluation Result

The multi-disease diagnosis task's model accuracy trend during exercise and authentication is displayed in the graph. The model remains knowledge patterns from the exercise data, as shown through the navy-blue line, which shows the training correctness, which increases gradually over epochs to about 70%. The proof accuracy, shown by the carroty line, stays comparatively stable at 50%, indicating that the model is not generalising healthy to new data [13]. When the perfect performs healthy on the training set but has trouble with fresh photos, this inconsistency among exercise and authentication accuracy is a sign of overfitting. Techniques like as regularisation, dropout tuning, learning rate adjustment, and data augmentation can be used to increase performance by improving generalisation and validation accuracy [Fig.4].

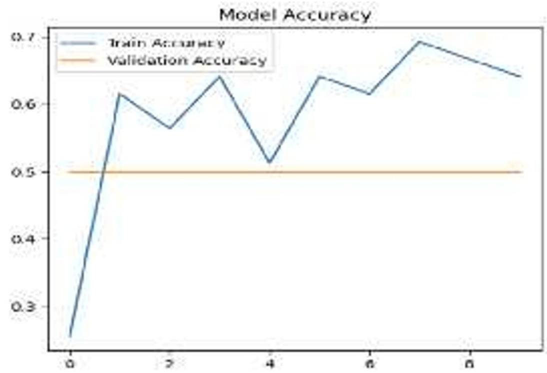

Fig.4 Model accuracy evaluation result

The multi-disease diagnosis model's training and validation loss trends are displayed in the graph. While the validation loss (orange line) stays larger and swings, indicating weaker generalisation to unknown data, the training loss (blue line) drops off dramatically, demonstrating good learning from the dataset. When the perfect achieves healthy on training data but badly on validation data, this trend points to overfitting. Model stability and generalisation can be enhanced with the use of strategies including data augmentation, regularisation, dropout, and early stopping [Fig.5].

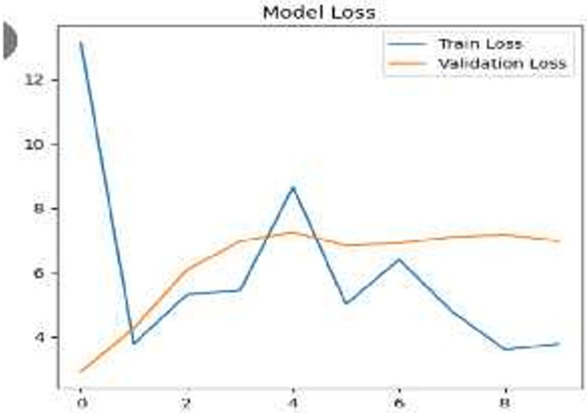

Fig.5 VGG19 train model loss

The model's capacity to differentiate between healthy and pathological chest X-ray pictures is evident from the prediction findings [4]. In the first instance, the uploaded X-ray image displays atypical white, hazy areas in the lung region, which are suggestive of inflammation and fluid buildup frequently linked to pneumonia [15]. These anomalous patterns are accurately recognised by the AI model, which then categorises the image as pneumonia. The second X-ray image, on the other hand, shows healthy lungs with normal heart and rib structure, as well as clear, evenly shaded lung fields free of aberrant opacities [14].
As a result, this scenario is correctly predicted by the model as Normal. These findings enhance initial diagnosis and medicinal decision-making by demonstrating the system's ability to analyse chest X-ray characteristics and provide accurate disease classification [Fig.6].

Fig.6 Prediction output result window

5 Conclusion

The suggested VGG19-based multi-disease diagnostic framework's performance evaluation shows partial convergence along with indications of model overfitting. The model matches the training data far better than unknown data, as evidenced by the training loss's constant decline over epochs while the validation loss stays rather high and varies. Similarly, the validation accuracy stays almost constant at 0.50, indicating weak generalization capability, but the training accuracy gradually increases from 0.25 to roughly 0.67. These results imply that the network's robustness on clinical test data has to be improved, even though it extracts discriminative visual characteristics during training. Future research should employ regularization techniques such dropout optimization, data augmentation, early stopping, and fine-tuning deeper VGG19 layers to enhance diagnostic generalization, lessen overfitting, and produce multi-disease prediction that is clinically reliable.

6 Acknowledgement

The authors would like to express their sincere gratitude to Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College (An Autonomous Institution), Chennai, for providing the academic support, technical resources, and research facilities necessary to carry out this research. The authors also extend their appreciation to the Department of Information Technology for their continuous encouragement, valuable guidance, and constructive feedback throughout the development of this work on Multi-Disease Diagnostic Framework Using VGG19-Based Deep Visual Feature Extraction from Medical Images.

References

1. Y. Xing, B. J. Meyer, M. Harandi, T. Drummond and Z. Ge, "Multimorbidity Content Based Medical Image Retrieval and Disease Recognition Using Multi Label Substitution Metric Learning," in *IEEE Access*, vol. **11**, pp. 50165-50179, (2023), doi: 10.1109/ACCESS.2023.3278376
2. S. K. Behera, K. Murali Gopal and S. B. Punuri, "A Deep Learning-Based Pneumonia Detection System with Explainable AI for Medical Decision Support," *2025 11th International Conference on Communication and Signal Processing (ICCSP)*, Melmaruvathur, India, (2025), pp. 694-699, doi: 10.1109/ICCSP64183.2025.11089357.
3. S. M. G, N. N. Das, N. Sunil, N. V, M. Rajeev and S. Gireesan, "Explainable AI for Disease Detection and Body Constitution Analysis Using Tongue Imaging: A Deep Learning Approach for Non-Invasive Diagnostics," *2025 2nd International Conference on ICTEST*, Ernakulam, India, (2025), pp. 1-6, doi: 10.1109/ICTEST64710.2025.11042469.
4. S. R. Chokhandre, P. Jadhav, N. Gole, A. N. Thakare, A. R. Bondade and P. Singh, "Clinical Breast Imaging and Deep Learning Model for Image Malignancy Prediction Using Resource-Constrained Embedded Devices," 2024 4th International Conference on Artificial Intelligence and Signal Processing (AISP), VIJAYAWADA, India, (2024), pp. 1-5, doi: 10.1109/AISP61711.2024.10870858.
5. C. Cui, H. Huo and T. Fang, "Deep Hashing With Multi-Central Ranking Loss for Multi-Label Image Retrieval," in *IEEE Signal Processing Letters*, vol. **30**, pp. 135-139, (2023), doi: 10.1109/LSP.2023.3244516.
6. I. Das *et al.*, "Improving Medical X-Ray Imaging Diagnosis with Attention Mechanisms and Robust Transfer Learning Techniques," in *IEEE Access*, vol. **13**, pp. 159002-159027, (2025), doi: 10.1109/ACCESS.2025.3607639.
7. S. Sangnark, P. Rattanachaisit, T. Patcharatrakul and P. Vateekul, "Explainable Multi-Modal Deep Learning with Cross-Modal Attention for Diagnosis of Dyssynergic Defecation Using Abdominal X-Ray Images and Symptom Questionnaire," in *IEEE Access*, vol. **12**, pp. 78132-78147, (2024), doi: 10.1109/ACCESS.2024.3409077.
8. V. -N. Pham, D. -T. Le, J. Bum, E. J. Lee, J. C. Han and H. Choo, "Attention-Aided Generative Learning for Multi-Scale Multi-Modal Fundus Image Translation," in *IEEE Access*, vol. **11**, pp. 51701-51711, (2023), doi: 10.1109/ACCESS.2023.3278596.
9. Y. Ba *et al.*, "High-Temperature Atomized Water–Assisted Reconstruction of $CsPbBr_3$ Wafers for High-Performance X-Ray Detection," in *IEEE Electron Device Letters*, vol. **45**, no. 7, pp. 1137-1140, July (2024), doi: 10.1109/LED.2024.3403668.
10. M. Niraula, M. M. Islam, Y. Nakashima, T. Matsubara, S. Hirano and Y. Takagi, "Growth and Characterization of Single Crystal Lead Halide Perovskites for X-Ray Detector Application," in *IEEE Transactions on Nuclear Science*, vol. **70**, no. 2, pp. 173-176, Feb. (2023), doi: 10.1109/TNS.2022.3230401.
11. S. Mandot et al., "A High-Sensitivity Benchtop X-Ray Fluorescence Emission Tomography (XFET) System with a Full-Ring of X-Ray Imaging-Spectrometers and a Compound-Eye Collimation Aperture," in IEEE Transactions on Medical Imaging, vol. **43**, no. 5, pp. 1782-1791, May (2024), doi: 10.1109/TMI.2023.3348791.
12. M. Pérez *et al.*, "Acquisition of Phase-Contrast X-Ray Images with Commercial Off-the-Shelf CMOS Image Sensors," in *IEEE Sensors Journal*, vol. **25**, no. 6, pp. 9618-9625, 15 March15, (2025), doi: 10.1109/JSEN.2025.3530784.
13. Q. Gao, H. Hu and W. Liu, "Traffic Sign Detection Under Adverse Environmental Conditions Based on CNN," in *IEEE Access*, vol. **12**, pp. 117572-117580, (2024), doi: 10.1109/ACCESS.2024.3446990.
14. X. Zhang, J. Chen, Y. Zhou, L. Han and J. Lin, "A Multiple-Layer Representation Learning Model for Network-Based Attack Detection," in *IEEE Access*, vol. **7**, pp. 91992-92008, (2019), doi: 10.1109/ACCESS.2019.2927465.
15. M. M. Safari, A. Sharifi, J. Mahmoodi and D. Abbasi-Moghadam, "Mesoscale Eddy Detection and Classification From Sea Surface Temperature Maps With Deep Neural Networks," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. **17**, pp. 10279-10290, (2024), doi: 10.1109/JSTARS.2024.3402823.