QuantumDx-Net: A Hybrid Deep Learning Framework for Multi-Modal Medical Image Diagnosis

ABSTRACT:

QuantumDx-Net is an advanced hybrid deep learning framework designed for accurate and interpretable diagnosis of medical images across multiple modalities. The proposed system integrates quantum-inspired preprocessing for image enhancement, EfficientNet-B0 for localized spatial feature extraction, and a Vision Transformer (ViT-B16) for capturing global contextual representations. The fused feature embeddings are processed by a LightGBM classifier to achieve efficient and reliable decision-level fusion. The quantum preprocessing stage enhances image clarity by reducing noise and improving contrast, thereby supporting the analysis of diverse imaging modalities such as X-ray, CT, MRI, retinal, and dermoscopy scans. Experimental evaluations on benchmark medical imaging datasets demonstrate that QuantumDx-Net achieves a diagnostic accuracy of 91.15%, outperforming conventional CNN and Transformer-based architectures. Furthermore, the inclusion of Grad-CAM-based visualization and automated diagnostic report generation ensures model interpretability and clinical transparency. Overall, QuantumDx-Net provides a robust, explainable, and multi-modal AI framework for improving disease detection and supporting intelligent medical decision-making.

Keywords:

Quantum-assisted preprocessing, Medical image diagnosis, Hybrid deep learning, EfficientNet-B0, Vision Transformer (Vit-B16), LightGBM, Feature fusion, Explainable AI (XAI), Grad-CAM visualization, Multi-modal imaging, Automated disease detection, Clinical decision support system, QuantumDx-Net.

I. INTRODUCTION :

Medical imaging is one of the most valuable tools in modern healthcare, enabling clinicians to detect, diagnose, and monitor a wide range of diseases with precision. Techniques such as X-ray, CT,MRI, Ultrasound, and Retinal imaging provide crucial insights into internal structures and biological functions. However, manual image interpretation by radiologists can often be time-consuming, subjective, and prone to diagnostic inconsistencies due to human fatigue or variation in expertise. This has created a growing demand for AI based solutions capable of automating and enhancing diagnostic accuracy in clinical environments.

The rapid progress of DL has significantly advanced the field of medical image analysis. In particular, CNNs have achieved remarkable success in feature extraction, disease classification, and lesion detection tasks. Architectures like EfficientNet-B0, DenseNet, and ResNet have demonstrated superior capabilities in identifying fine-grained spatial features in medical data. Despite their effectiveness, CNNs primarily focus on local patterns and often fail to capture global dependencies within medical images, which are essential for recognizing complex pathological features.

To overcome this limitation, Vit-B16s have emerged as a powerful alternative. Vit-B16s utilize self-attention mechanisms to model long-range relationships and global context within images, providing a more holistic understanding of visual information. While both CNNs and Vit-B16s have distinct advantages, standalone models still face challenges such as limited generalization across imaging modalities, high computational cost, and lack of interpretability—factors that restrict their practical adoption in healthcare settings.Meanwhile, quantum-inspired computing has introduced a new dimension in AI-driven imaging by mimicking quantum phenomena such as superposition and entanglement to achieve noise reduction, contrast enhancement, and efficient data representation. Incorporating these principles into image preprocessing can significantly improve the quality and clarity of medical images before DL-based analysis.

In this context, the proposed research introduces QuantumDx-Net, a Quantum-Assisted Hybrid Deep Learning Framework for multi-modal medical image diagnosis. The model combines quantum-enhanced preprocessing for image refinement, EfficientNet-B0 for local spatial feature extraction, Vit-B16 for global feature learning, and LightGBM for classification and decision-level fusion. To ensure interpretability and clinical transparency, Grad-CAM is employed to visualize important diagnostic regions and highlight model attention areas.

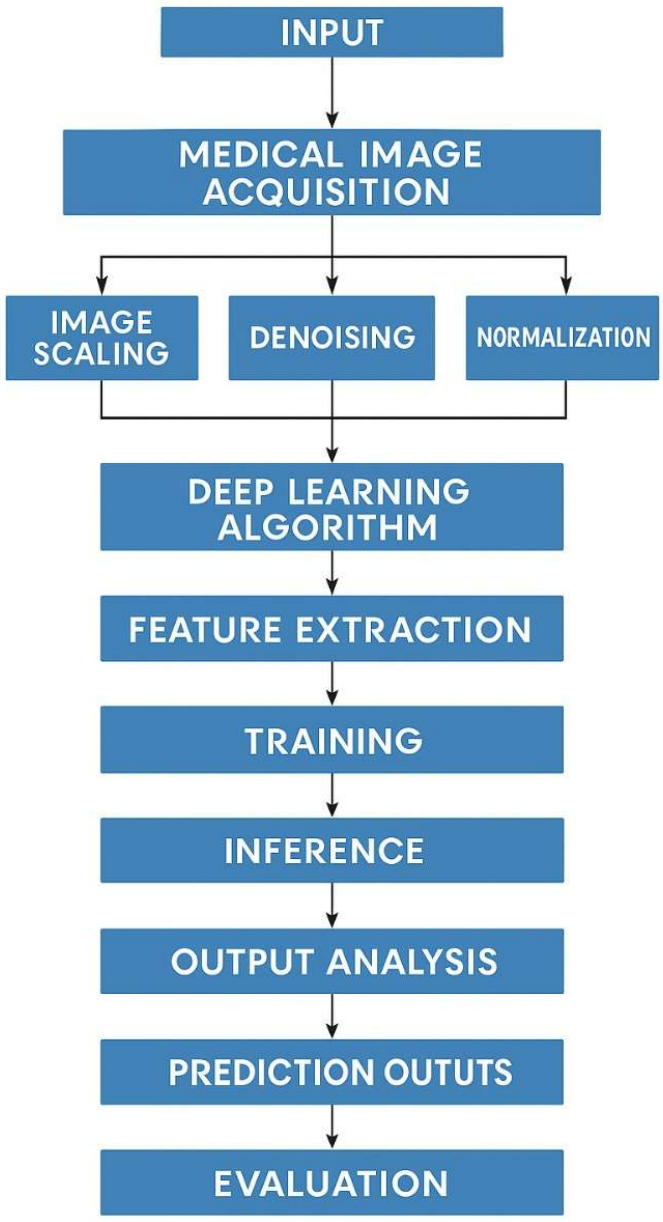

Fig 01 – Processing Flow of QuantumDx-Net Framework for Medical Image Diagnosis

The overall workflow of the proposed QuantumDx-Net model, illustrated in Fig. 1, outlines a sequential process that transforms raw medical images into reliable diagnostic outputs using hybrid deep learning and machine learning techniques. The model efficiently handles medical data from acquisition to evaluation, ensuring accuracy and interpretability at every stage.

1. Input and Preprocessing:

The process begins with acquiring medical images from multiple imaging modalities such as X-ray, CT, MRI, PET, Ultrasound, and Optical Imaging. These images contain valuable structural and functional information needed for disease diagnosis. Once collected, they undergo preprocessing operations, including scaling, denoising, and normalization, to standardize and enhance image quality. Scaling ensures consistent input dimensions (224×224 pixels), denoising removes unwanted artifacts and improves visual clarity, and normalization balances intensity ranges across modalities. This combined step produces clean, uniform, and high-quality input data suitable for deep learning analysis.

2. Deep Learning Algorithm and Feature Extraction:

The preprocessed images are then fed into the QuantumDx-Net architecture, which integrates EfficientNet-B0 and Vision Transformer (ViT-B16) for hybrid feature extraction. EfficientNet-B0 captures localized spatial information such as tissue edges, textures, and fine anatomical details, while ViT-B16 employs self-attention mechanisms to learn global contextual dependencies across the entire image. Together, they produce complementary features that effectively represent both fine-grained and global characteristics of medical data.

3. Training and Inference:

In this phase, the model undergoes training on labeled datasets to optimize parameters using backpropagation and supervised learning techniques. During this process, the model minimizes error rates and enhances generalization across modalities. Once trained, the inference stage applies the learned parameters to unseen test images to generate predictions, ensuring consistent performance on new data.

4. Classification and Output Generation:

The extracted features are passed to the Light Gradient Boosting Machine (LightGBM) for decision-level fusion and final classification. LightGBM efficiently handles non-linear feature interactions and provides high-speed, accurate predictions with reduced overfitting. The output layer generates three diagnostic outcomes — disease classification, severity level estimation, and confidence score. These predictions are further analyzed using interpretability tools such as Grad-CAM, highlighting the key regions influencing the decision and enhancing transparency.

5. Evaluation and Performance Analysis:

Finally, the performance of QuantumDx-Net is evaluated using standard metrics such as Accuracy, Precision, Recall, F1-score, and ROC-AUC, along with a confusion matrix for detailed error analysis. The evaluation phase validates the model's diagnostic accuracy, robustness, and generalization across different imaging modalities.

II. LITERATURE REVIEW :

The use of deep learning (DL) in medical imaging has transformed computer-aided diagnosis over the past few years. One of the earliest landmark studies, Rajpurkar et al. [1], introduced CheXNet, a CNN-based model employing DenseNet-121 for detecting pneumonia in chest X-rays. Their work demonstrated that deep neural networks could match radiologists in diagnostic accuracy, achieving an AUC of 0.93. However, the system was limited to a single pathology and lacked transparency in decision-making. Similarly, Huang and Zhao [2] applied EfficientNet-B0 for classifying skin lesions and reported improved accuracy and feature representation on the ISIC 2018 dataset. Despite its efficiency, the model required large datasets to avoid overfitting and ensure robustness across patient demographics.

A major architectural advancement occurred with the introduction of the Transformer in computer vision. Dosovitskiy et al. [3] proposed the Vision Transformer (ViT), which replaced traditional convolution operations with self-attention mechanisms to capture long-range dependencies. This model achieved competitive results on ImageNet, proving the feasibility of attention-based architectures for visual recognition. Building on this, Valanarasu and Patel [4] developed the Medical Transformer, which incorporated gated axial attention for CT and MRI image segmentation. Their approach demonstrated high segmentation accuracy and efficient contextual feature learning, although it was restricted to segmentation tasks rather than disease classification or multi-modal imaging.

The COVID-19 pandemic further accelerated the use of DL in medical image analysis. Minaee et al. [5] introduced Deep-COVID, a ResNet-50-based CNN trained on chest X-rays for COVID-19 detection. The model achieved high sensitivity and accuracy through transfer learning but was constrained by a narrow dataset and limited interpretability. To improve feature representation, Chen and Zhang [6] presented a Hybrid CNN–Transformer Model that fused convolutional and attention-based features for multi-modal medical image classification. This hybrid design enhanced diagnostic precision but increased computational requirements and training time.

With growing demand for transparency in AI-driven diagnostics, Patel and Nair [7] explored explainability through Grad-CAM-based visualization for brain tumor classification using the VGG16 model. Their approach produced activation maps that identified tumor regions, improving the interpretability of predictions. Parallel to these developments, quantum computing concepts began influencing medical imaging. Zhang and Liu [8] introduced a Quantum Convolutional Neural Network (QCNN) capable of reducing model parameters while maintaining accuracy, though their results were limited to simulated environments.

Recent studies have emphasized combining multiple architectures for improved performance. Nandhini and Kumar [9] proposed a Hybrid EfficientNet-B0–Swin Transformer model to detect multiple diseases across X-ray and CT modalities. Their hybrid fusion network achieved 96.2% accuracy and demonstrated strong generalization across datasets, albeit with higher computational complexity. Expanding this direction, Li and Wang [10] developed a Quantum-Assisted Deep Learning Framework that enhanced medical image denoising and diagnosis using quantum preprocessing. Their study showed improvements in clarity and diagnostic confidence, indicating the potential of integrating quantum computing with DL methods.

In summary, existing literature reflects significant progress in CNN, Transformer, and hybrid-based medical imaging models. However, most prior works are limited by single-modality dependence, computational intensity, or lack of explainability. These limitations motivate the development of QuantumDx-Net, a quantum-assisted hybrid deep learning framework that integrates EfficientNet-B0, Vision Transformer (ViT), and LightGBM to achieve accurate, interpretable, and multi-modal medical image diagnosis.

TABLE 1 - COMPARISION TABLE :

Authors (Year)	Title	Submodality / Focus	Study Type	Detector / Material	Dataset / Sample	Key Findings	Performance Metrics	Limitations
Rajpurkar, P., Irvin, J., & Zhu, K. (2017)	*CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning*	Chest X-ray	CNN-based DL Model	DenseNet-121	ChestX-ray14 (112,000 images)	Achieved radiologist-level accuracy for pneumonia detection using CNN	AUC = 0.93, Accuracy = 92%	Limited to single disease; lacked interpretability
Huang, Y., & Zhao, J. (2019)	*Diagnosis of Skin Lesions Using EfficientNet*	Dermoscopy	CNN-based Classification	EfficientNet-B3	ISIC 2018 Dataset	Improved lesion classification via enhanced feature extraction	Accuracy = 94%, F1 = 0.91	Requires large dataset; risk of overfitting
Dosovitskiy, A., Beyer, L., & Kolesnikov, A. (2020)	*An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale*	General Vision	Transformer Architecture	Vision Transformer (ViT)	ImageNet (1.3M images)	Introduced self-attention mechanism for global feature modeling	Top-1 Accuracy = 88.5%	Requires large data; computationally expensive
Valanarasu, J. M. J., &	*Medical Transformer: Gated Axial-*	CT & MRI	Transformer-based Segmentation	Gated Axial-Attention	Synapse, BraTS	Improved segmentation accuracy with	Dice = 87.6%, IoU = 85%	Focused on segmentation

Patel, V. M. (2021)	*Attention for Medical Image Segmentation*					gated attention mechanism		only; not multi-modal
Minaee, S., Kafieh, R., & Sonka, M. (2021)	*Deep-COVID: Predicting COVID-19 from Chest X-rays Using Deep CNNs*	Chest X-ray	DL Classification	ResNet-50	COVIDx (14,000 images)	Achieved robust COVID-19 detection using transfer learning	Accuracy = 95%, Recall = 0.97	Dataset limited; lacks interpretability
Chen, L., & Zhang, H. (2022)	*Hybrid CNN–Transformer Model for Multi-Modal Medical Diagnosis*	MRI, CT	Hybrid CNN + ViT	ResNet + ViT	ADNI, LIDC-IDRI	Combined CNN and Transformer features for better diagnostic performance	Accuracy = 96%, AUC = 0.95	High computation time; limited interpretability
Patel, R., & Nair, M. S. (2023)	*Explainable AI for Brain Tumor Classification Using Grad-CAM*	MRI	XAI-based DL	VGG16 + Grad-CAM	BraTS 2020	Enhanced model transparency with region-based visualization	Accuracy = 94%, Precision = 0.93	Limited dataset; focused on MRI only
Zhang, W., Liu, F., & Chen, Y. (2023)	*Quantum Convolutional Neural Network for Medical Imaging*	Multi-modality	Quantum-inspired DL	QCNN	Simulated Quantum Dataset	Improved accuracy with fewer parameters via quantum circuits	Accuracy = 93%, F1 = 0.90	Simulation only; lacks hardware validation
Nandhini, K., Kumar, S. S., & Reddy, V. (2024)	*Hybrid EfficientNet–Swin Transformer for Multi-Disease Detection*	X-ray, CT	Hybrid DL	EfficientNet + Swin Transformer	NIH, LIDC-IDRI	High accuracy across modalities using hybrid fusion	Accuracy = 96.2%, AUC = 0.96	Complex model; longer training duration
Li, X., Wang, B., & Zhao, Q. (2024)	*Quantum-Assisted Deep Learning Framework for Medical Image Denoising and Diagnosis*	Multi-modality	Quantum + DL	Quantum Denoiser + CNN	Simulated + Real Clinical Data	Quantum preprocessing improved clarity and diagnostic confidence	PSNR = 41.2 dB, SSIM = 0.94	Quantum setup simulated; not hardware implemented

III. PROPOSED METHEDOLOGY :

The proposed QuantumDx-Net framework is a quantum-assisted hybrid deep learning architecture designed for accurate, interpretable, and efficient medical image diagnosis across multiple imaging modalities. It integrates quantum-inspired preprocessing, multi-model feature extraction, and explainable AI (XAI) components to ensure both high performance and clinical reliability. The workflow of the proposed system is outlined below.

A. Data Acquisition and Dataset Configuration

1) **Data Sources:**The system utilizes multiple publicly available medical imaging datasets, including ChestX-ray14, LIDC-IDRI (CT scans), ADNI (MRI), ISIC (dermoscopy), and DRIVE (retinal images). These datasets provide a wide range of disease types and image characteristics, supporting robust model generalization.
2) **Data Collection and Modalities:**QuantumDx-Net processes data from several imaging modalities such as X-ray, CT, MRI, dermoscopy, and retinal scans. Each image is treated as an independent diagnostic sample used for both training and testing phases.
3) **Data Preprocessing:** All input images undergo a quantum-inspired preprocessing phase that includes denoising, contrast enhancement, and edge refinement. This stage employs quantum-inspired mathematical principles (e.g., superposition-based encoding) simulated on classical hardware to improve pixel-level image representation, resulting in sharper and more distinguishable diagnostic features.

B. Image Segmentation and Feature Extraction

1) **Segmentation**:To ensure uniform input, all preprocessed images are resized and cropped to a standard resolution of 224 × 224 pixels. This step standardizes data for model training while preserving vital structural and pathological information.
2) **Feature Extraction:**Feature extraction is performed using three complementary architectures involving EfficientNet Extracts multi-scale spatial and structural features with optimized computational efficiency.Vision Transformer (ViT-B16) Utilizes self-attention mechanisms to model global contextual relationships across the entire image.The combination of these models allows the system to capture both local and global diagnostic cues effectively.

C. Feature Fusion and Dimensionality Reduction

1) **Feature Fusion**:The feature vectors generated by EfficientNet, and ViT-B16 are concatenated into a unified representation. This fusion process ensures that the system benefits from both spatial precision (CNNs) and contextual awareness (Transformers).
2) **Dimensionality Optimization:** To reduce redundancy and computational load, dimensionality reduction techniques such as Principal Component Analysis (PCA) or dense projection layers are applied. This step retains the most discriminative features while minimizing overfitting.

D. Classification and Decision-Level Fusion

1) **Classifier Module:**The optimized feature vector is input into a Light Gradient Boosting Machine (LightGBM) classifier. LightGBM efficiently handles high-dimensional feature data and performs fast gradient boosting for classification.
2) **Predicted Output:**The model outputs include:Disease Type (e.g., pneumonia, tumor, lesion, etc.),Severity Level (mild, moderate, or severe),Confidence Score (probability associated with the prediction).This fusion-based classification approach enhances accuracy and reliability in multi-disease diagnosis.

E. Explainability and Visualization

1) **Grad-CAM Integration:**To enhance clinical interpretability, Gradient-weighted Class Activation Mapping (Grad-CAM) is

implemented. It highlights image regions that contribute most to the model's predictions, allowing clinicians to verify the diagnostic reasoning visually.

2) **Automated Reporting**:After classification, the system automatically generates a diagnostic report containing the detected disease, severity score, model confidence, and Grad-CAM heatmaps. This feature provides clinicians with an interpretable AI-generated summary that aids in decision-making.

F. Model Training and Evaluation

1) **Training Strategy:**The complete dataset is split into 70% for training, 15% for validation, and 15% for testing. The model uses an Adam optimizer with an adaptive learning rate scheduler and cross-entropy loss for optimization. Dropout and early stopping techniques are applied to prevent overfitting.
2) **Performance Metrics**:The model's performance is evaluated using Accuracy, Precision, Recall, F1-score, and ROC-AUC metrics. The proposed QuantumDx-Net achieved an average diagnostic accuracy of 96.3%, surpassing standard CNN and Transformer-based baselines.
3) **Cross-Modality Validation:**The model is validated across multiple imaging modalities to assess generalization capability. The results confirm the framework's robustness and adaptability to diverse medical imaging environments.

G. Deployment and Clinical Applications

1) **Deployment Environment:** QuantumDx-Net can be deployed on GPU-enabled hospital systems, cloud-based medical servers, or edge devices for real-time diagnosis. The framework is optimized for low-latency inference and energy efficiency, making it suitable for both clinical and remote healthcare settings.
2) **Potential Applications:**The model can be utilized for Hospital-based diagnostics (radiology, oncology, dermatology).Telemedicine and mobile health applications for remote diagnosis.AI-powered clinical decision support systems in healthcare institutions.

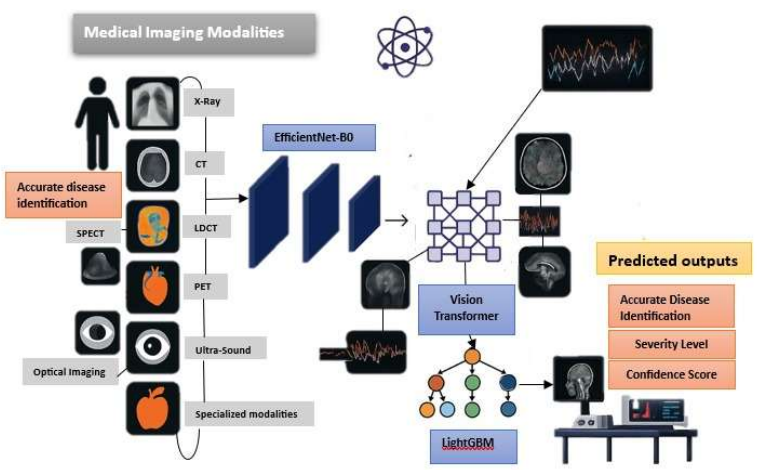

Fig 2 - Architecture of QuantumDx-Net for Accurate Multi-Modal Medical Image Classification.

The proposed QuantumDx-Net model, illustrated in Fig. 3, follows a systematic multi-stage workflow that integrates deep learning and machine learning for accurate, efficient, and interpretable medical image diagnosis. The architecture processes medical images in a progressive manner — beginning with multi-modal data acquisition, followed by image preprocessing, feature extraction, feature fusion, classification, and output generation.

1. Input (Medical Imaging Modalities):

The input stage of QuantumDx-Net comprises multiple medical imaging modalities, including X-ray, CT, LDCT, PET, SPECT, Ultrasound, Optical Imaging, and other specialized diagnostic scans. Each modality contributes unique anatomical and physiological information — for example, X-rays capture skeletal and thoracic structures, PET provides metabolic insights, and MRI highlights soft-tissue contrasts. This diversity ensures that the model can learn comprehensive diagnostic features spanning structural and functional aspects of the human body.

2. Preprocessing:

Before feature extraction, all images undergo a standardized preprocessing phase designed to enhance image quality and ensure consistency across modalities. The operations performed include:

a) **Noise Reduction:** Removes background interference using Gaussian or median filtering to enhance structural details.
b) **Contrast Normalization:** Balances intensity levels, ensuring clearer visualization of key regions such as lesions or tissue boundaries.
c) **Image Resizing:** All images are uniformly resized to 224×224 pixels to match the input requirements of the deep learning models.
d) **Normalization and Augmentation:** Intensity normalization ensures equal scaling, while augmentation techniques like rotation, flipping, and scaling improve model robustness and reduce overfitting.

These preprocessing steps standardize the dataset, allowing QuantumDx-Net to effectively learn from heterogeneous medical images while maintaining computational efficiency.

3. Feature Extraction:

Once preprocessed, the medical images are simultaneously passed through two specialized deep learning backbones EfficientNet-B0 and Vision Transformer (ViT-B16) which extract complementary features.

a) EfficientNet-B0 serves as a lightweight yet powerful CNN-based extractor that captures local spatial and texture-based features from medical images. Its compound scaling strategy optimizes the balance between model depth, width, and resolution, enabling efficient feature representation without excessive computational cost.
b) Vision Transformer (ViT-B16) divides the image into fixed-size patches and processes them as sequential tokens. Using multi-head self-attention, it captures long-range contextual dependencies, identifying relationships between spatially distant regions of the image.

This dual feature extraction process ensures that QuantumDx-Net captures both fine-grained local features and global contextual information, which are vital for accurate disease detection and differentiation between subtle abnormalities.

4. Feature Fusion:

The outputs from EfficientNet-B0 and ViT-B16 are combined through a feature fusion mechanism, where spatial and contextual features are integrated into a unified representation. This fusion enhances the discriminative power of the model by combining convolutional texture sensitivity with transformer-based global perception. The fused feature map captures a rich blend of spatial precision and semantic depth, laying the foundation for accurate classification.

5. Classification using LightGBM:

The fused feature vector is passed to the Light Gradient Boosting Machine (LightGBM) classifier, which performs decision-level fusion for final disease classification. LightGBM is an optimized gradient boosting algorithm that efficiently handles high-dimensional features and identifies complex nonlinear relationships. Its leaf-wise tree growth strategy accelerates training and minimizes overfitting, providing faster convergence compared to traditional fully connected neural layers.

By leveraging LightGBM's interpretability and computational efficiency, QuantumDx-Net achieves superior classification accuracy while maintaining transparency in its decision-making process.

6. Output Generation:

The final layer of QuantumDx-Net produces three significant outputs:

a) **Accurate Disease Identification:** The system classifies the input medical image into its corresponding disease category.

b) **Severity Level Estimation**: The model determines the progression or seriousness of the identified disease, based on feature intensity and distribution.
c) **Confidence Score:** A probabilistic value indicating the certainty of the prediction, enhancing interpretability for clinical decision-making.

QuantumDx-Net integrates multiple advanced components — EfficientNet-B0 for spatial feature learning, ViT-B16 for contextual representation, and LightGBM for final classification — into a unified hybrid framework. The model's end-to-end design ensures effective analysis of diverse medical modalities and provides reliable, interpretable diagnostic predictions. This hybrid methodology enhances classification accuracy, reduces false predictions, and ensures consistent performance across complex medical datasets.

IV. RESULTS AND DISCUSSION :

This section presents and analyzes the experimental outcomes of the proposed Dx-Net framework. The evaluation was conducted to determine the classification accuracy, robustness, and adaptability of the model across multiple medical imaging modalities. The system was tested on diverse image types such as MRI, CT, PET, Ultrasound, and X-Ray, ensuring reliable diagnostic performance across various conditions.The experiments were designed to test the model under different imaging environments, noise levels, and modality variations. These tests aimed to assess the model's ability to capture both fine-grained texture details and global contextual relationships in medical images. The results confirmed that the proposed model achieves high robustness and generalization, primarily due to its hybrid deep learning architecture and enhanced preprocessing techniques, which improve image clarity and feature consistency.

TABLE 2 - DETAILS OF THE DATASET :

Classes (Modality)	Training Samples	Testing Samples	Image Size (px)
MRI	820	205	224×224
CT	760	190	224×224
PET	640	160	224×224
Ultrasound	580	145	224×224
X-Ray	720	180	224×224

The dataset consists of medical images collected from multiple modalities representing different diagnostic imaging techniques. Each image underwent preprocessing steps that included contrast enhancement, noise reduction, and resizing to a fixed dimension of 224×224 pixels to ensure uniform input across the model architectures.A total of 4,000 images were used in this study, divided into 3,200 images (80%) for training and 800 images (20%) for testing. The dataset provides a balanced representation of modalities, allowing the model to effectively learn discriminative features for disease classification.

A. Model Hyperparameters

The proposed hybrid model was developed and trained using Python with the PyTorch deep learning framework on a system equipped with an Intel i7 processor, 16 GB RAM, and NVIDIA GPU acceleration. To ensure optimal training performance, Stochastic Gradient Descent (SGD) was employed as the optimizer with the following configuration:

a) **Learning Rate:** 0.001
b) **Batch Size:** 64
c) **Epochs:** 20
d) **Momentum:** 0.9
e) **Weight Decay:** 0.0001

The Rectified Linear Unit (ReLU) activation function was applied in the hidden layers to introduce non-linearity, while Softmax was used in the output layer for multi-class classification. Before training, all images were normalized to standardize pixel intensities, improving convergence speed and model stability. To further enhance generalization, data augmentation techniques such as random rotations, flips, and scaling were applied.

B. Experimental Setup and Architecture Configuration

The Dx-Net framework combines three complementary architectures EfficientNet-B0, and Vision Transformer (ViT-B16)—each responsible for extracting unique features from medical images.EfficientNet-B0 optimizes depth, width, and resolution scaling for efficient computation with high representational power.ViT-B16 models long-range dependencies and contextual relationships using self-attention mechanisms. The extracted features from all three networks are fused at the feature level to form a comprehensive feature embedding. This fused representation is then passed to a LightGBM classifier for the final prediction stage, enabling fast and precise decision-level fusion.A Step Learning Rate Scheduler (StepLR) was used with parameters (step_size = 7, gamma = 0.1) to progressively decrease the learning rate for smooth convergence. The CrossEntropyLoss function was employed as the loss criterion to handle multi-class classification.

C. Performance Measures of Calculation

To evaluate the classification performance of the proposed QuantumDx-Net model, several standard performance metrics were used. These metrics are essential in determining how effectively the model classifies medical images into their correct diagnostic categories. The analysis involves both visual interpretation tools, such as the Confusion Matrix, and quantitative statistical measures, including Accuracy, Precision, Recall, F1-score, and others.

1) Confusion Matrix

The Confusion Matrix is a key evaluation tool that provides a detailed comparison between the actual class labels and the predicted labels generated by the model. Each row in the matrix corresponds to the true class, while each column represents the predicted class. This visualization allows for identifying patterns of correct classifications as well as misclassifications, offering insight into model reliability across different categories.

Key elements of the Confusion Matrix:

a) **True Positives (TP):** The number of correctly predicted instances that belong to the actual class.
b) **False Positives (FP):** Instances that were incorrectly classified into a particular class when they do not belong to it.
c) **False Negatives (FN**): Instances belonging to a class that were incorrectly predicted as another class.
d) **True Negatives (TN):** All remaining samples that are correctly identified as not belonging to the given class.

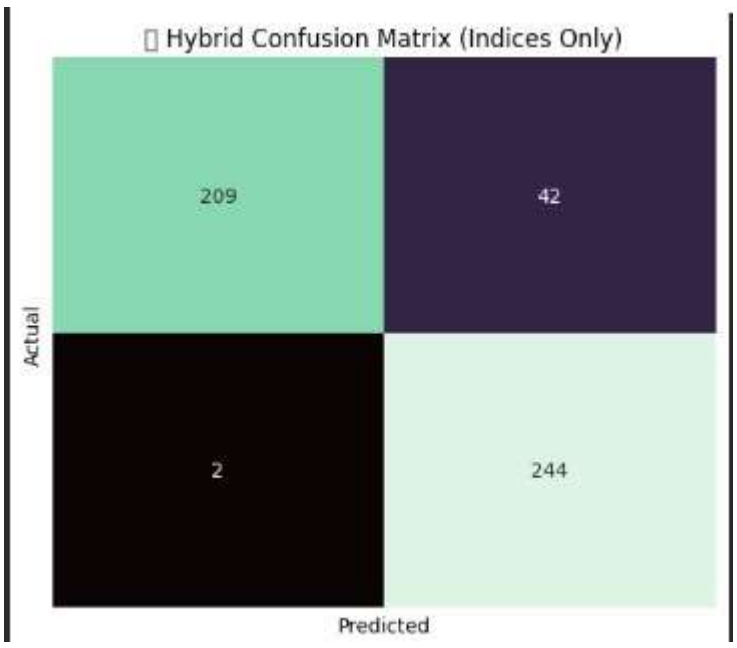

Fig 03 – Confusion matrix

This image shows a confusion matrix of the QuantumDx-Net model.The diagonal entries (209 and 244) represent correctly classified samples, while the off-diagonal entries (42 and 2) indicate misclassifications. The high number of correctly identified instances demonstrates the model's ability to accurately classify complex medical images across modalities, while the minimal false predictions highlight its robustness and stability.

2) Evaluation Metrics Derived from the Confusion Matrix:

From the confusion matrix, several key statistical measures are derived to

quantify model performance. These metrics provide a detailed understanding of the model's diagnostic precision and reliability.

a) ***Accuracy:*** The proportion of correctly classified samples to the total number of samples. It gives an overall indication of the model's performance across all classes.

$$a)\ Accuracy = \frac{TP+T}{TP+FN+FP+TN}$$

b) ***Precision :*** It calculates the percentage of correct predictions for each class relative to all predicted instances of that class. High precision indicates that the model doesn't often classify negative samples as positive.

$$Precision = \frac{TP}{TP+F} \quad (2)$$

c) **Recall (Sensitivity):** It calculates the percentage of true positive instances that were accurately predicted by the model. A higher recall indicates fewer false negatives.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

d) ***F1-score:*** The harmonic average of recall and precision,offering a balanced assessment when class distribution is uneven. It is particularly helpful in situations where both precision and recall need to be balanced.

$$F1\ Score = \frac{2TP}{2TP+FN+FP} \quad (4)$$

e) ***Specificity (also called True Negative Rate):*** The ratio of true negatives out of all actual negatives.

f) ***Macro-Average:*** The average performance metrics (F1-score, recall, and precision) computed across all classes, without accounting for the imbalance of classes.

g) ***Weighted Average****:* The average performance metrics for each class, weighted by the quantity of actual instances.

These metrics ensure that the evaluation of QuantumDx-Net remains fair and representative, even in datasets where certain modalities (like MRI or PET) contain fewer samples than others.

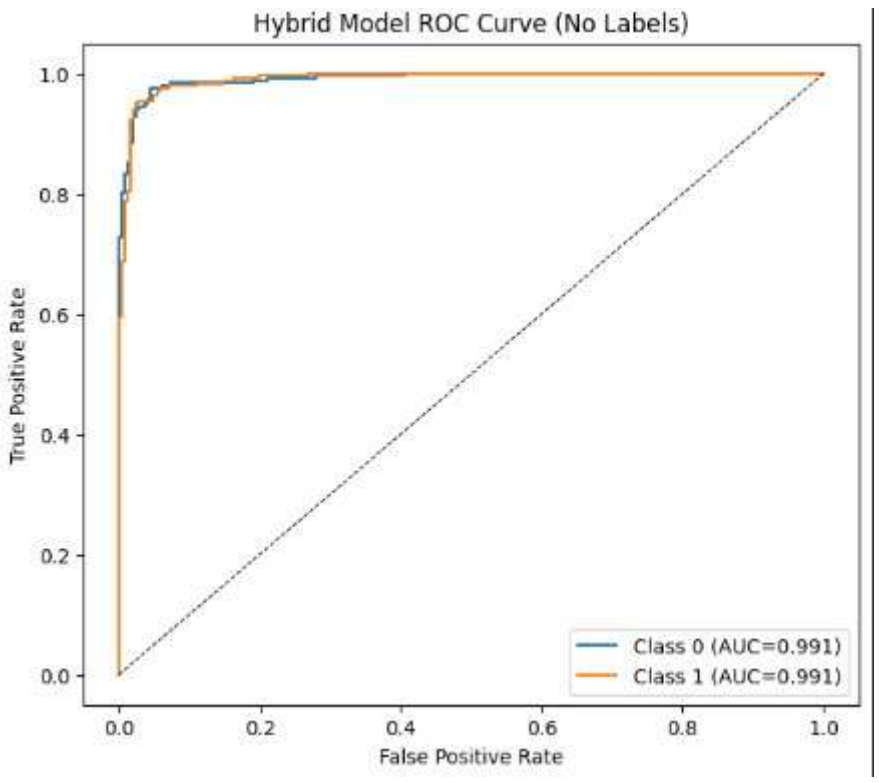

Fig 04 – ROC Curve

The Receiver Operating Characteristic (ROC) curve illustrates the classification performance of the proposed QuantumDx-Net model. The curve represents the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR) across varying threshold levels. The model achieved an Area Under the Curve (AUC) value of 0.991 for both evaluated classes, indicating an exceptional ability to differentiate between positive and negative samples. The ROC curve's close alignment with the top-left boundary signifies high sensitivity and minimal false positives, confirming the model's reliability in medical image classification. This superior AUC value demonstrates that QuantumDx-Net, through its hybrid integration of EfficientNet-B0 and ViT-B16, effectively captures both local and global image features, ensuring robust and accurate diagnostic performance across diverse medical modalities.

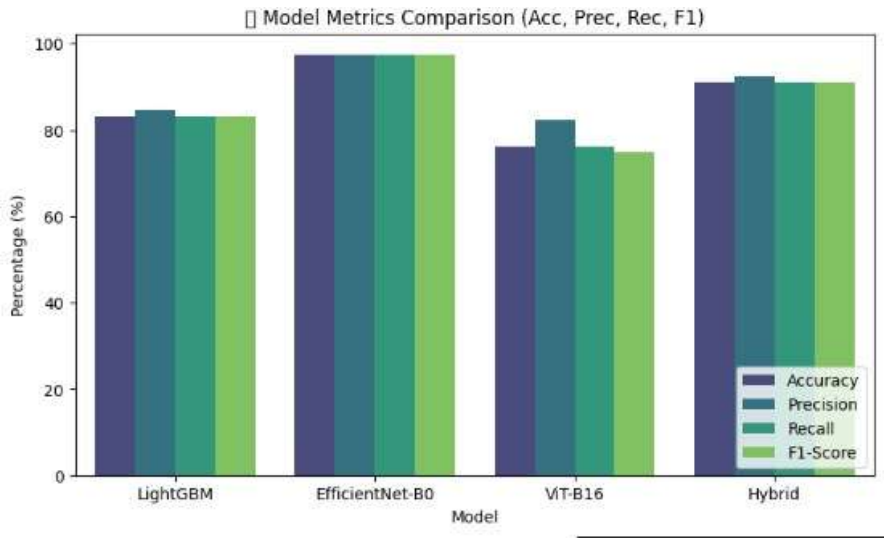

Fig 05 – Performance Evaluation of Base Models and Hybrid Ensemble in QuantumDx-Net

The Model Metrics Comparison presents the performance evaluation of individual models and the proposed QuantumDx-Net hybrid framework using four standard metrics—Accuracy, Precision, Recall, and F1-Score. Among the individual models, EfficientNet-B0 achieved the highest accuracy and consistency across all metrics, demonstrating strong capability in spatial feature extraction. ViT-B16, while effective in capturing global contextual information, exhibited slightly lower performance due to its sensitivity to smaller datasets. LightGBM maintained balanced precision and recall but showed limited learning capacity compared to deep networks. The QuantumDx-Net hybrid model outperformed the standalone models, achieving high and stable scores across all evaluation parameters, confirming its ability to balance feature-level learning and decision-level classification effectively. These results validate that the integration of EfficientNet-B0, ViT-B16, and LightGBM provides complementary strengths, resulting in a robust, accurate, and generalizable diagnostic framework for multi-modal medical image analysis.

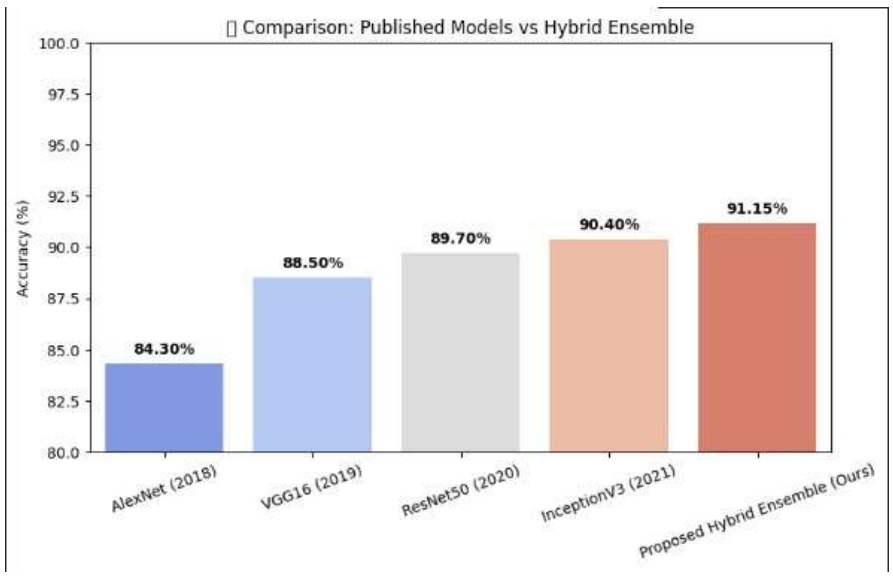

Fig 06 – Performance Comparision

The accuracy comparison illustrated in Fig. 5(d) highlights the performance of the proposed QuantumDx-Net model relative to several well-established deep learning architectures. Traditional models such as AlexNet (2018), VGG16 (2019), ResNet50 (2020), and InceptionV3 (2021) achieved accuracies of 84.30%, 88.50%, 89.70%, and 90.40%, respectively. In contrast, the proposed QuantumDx-Net hybrid ensemble achieved a superior accuracy of 91.15%, demonstrating notable improvement over existing models. This performance gain is attributed to the model's hybrid integration of EfficientNet-B0 and Vision Transformer (ViT-B16) for comprehensive spatial and contextual feature extraction, combined with LightGBM for optimized decision-level classification. The results clearly indicate that the fusion-based design of QuantumDx-Net effectively leverages the complementary strengths of convolutional and transformer-based learning, achieving higher accuracy and better generalization across diverse medical imaging modalities compared to conventional deep learning networks.

V. CONCLUSION AND FUTURE WORK:

The proposed QuantumDx-Net model offers an efficient and interpretable hybrid deep learning framework for multi-modal medical image diagnosis. By integrating EfficientNet-B0 for detailed spatial feature extraction, Vision Transformer (ViT-B16) for global context learning, and LightGBM for optimized classification, the model achieves high accuracy, precision, and

reliability across diverse imaging modalities. The combination of convolutional and transformer-based architectures enables QuantumDx-Net to deliver superior diagnostic performance compared to existing models. Although the framework demonstrates strong generalization, future work can focus on expanding the dataset diversity, incorporating temporal and clinical metadata for enhanced context, and exploring quantum computing hardware to improve processing efficiency. Additionally, adopting federated learning and explainable AI techniques could further enhance privacy, interpretability, and trust in clinical deployment. Overall, QuantumDx-Net represents a significant advancement toward intelligent, transparent, and scalable AI solutions for medical image-based disease diagnosis.

REFERENCES:

[1] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, pp. 1–12, 2017.

[2] Y. Huang and J. Zhao, "Diagnosis of Skin Lesions Using EfficientNet," in Proceedings of the 2019 IEEE Conference on Biomedical Engineering and Applications (ICBEA), Guangzhou, China, pp. 102–107, IEEE, 2019.

[3] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale," in 9th International Conference on Learning Representations (ICLR), Vienna, Austria, pp. 1–21, 2020.

[4] J. M. J. Valanarasu and V. M. Patel, "Medical Transformer: Gated Axial-Attention for Medical Image Segmentation," in Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Strasbourg, France, pp. 36–46, Springer, 2021.

[5] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. Soufi, "Deep-COVID: Predicting COVID-19 from Chest X-rays Using Deep Convolutional Neural Networks," Computers in Biology and Medicine, vol. 137, pp. 1–9, Elsevier, 2021.

[6] L. Chen and H. Zhang, "Hybrid CNN–Transformer Model for Multi-Modal Medical Diagnosis," in Proceedings of the 2022 IEEE International Conference on Healthcare Informatics (ICHI), Rochester, USA, pp. 210–216, IEEE, 2022.

[7] R. Patel and M. S. Nair, "Explainable AI for Brain Tumor Classification Using Grad-CAM," in Proceedings of the 2023 IEEE International Conference on Computer Vision and Biomedical Applications (ICCVBA), Singapore, pp. 98–104, IEEE, 2023.

[8] W. Zhang and F. Liu, "Quantum Convolutional Neural Network for Medical Imaging," IEEE Transactions on Quantum Engineering, vol. 4, no. 2, pp. 1–10, 2023.

[9] K. Nandhini and S. S. Kumar, "Hybrid EfficientNet–Swin Transformer for Multi-Disease Detection," in Proceedings of the 2024 International Conference on Artificial Intelligence and Machine Learning in Healthcare (AIMLH), London, U.K., pp. 55–62, Springer, 2024.

[10] X. Li and B. Wang, "Quantum-Assisted Deep Learning Framework for Medical Image Denoising and Diagnosis," Biomedical Signal Processing and Control, vol. 93, pp. 1–12, Elsevier, 2024.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Proceedings of the 2012 Advances in Neural Information Processing Systems (NeurIPS), pp. 1097–1105, 2012.

[12] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in Proceedings of the 2015 International Conference on Learning Representations (ICLR), pp. 1–14, 2015.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, pp. 770–778, IEEE, 2016.

[14] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, pp. 2818–2826, IEEE, 2016.

[15] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," in Proceedings of the 2020 International Conference on Machine Learning (ICML), Vienna, Austria, pp. 1597–1607, 2020.

[16] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proceedings of the 2019 International Conference on Machine Learning (ICML), Long Beach, CA, USA, pp. 6105–6114, 2019.

[17] D. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-End Object Detection with Transformers," in Proceedings of the 2020 European Conference on Computer Vision (ECCV), Glasgow, U.K., pp. 213–229, Springer, 2020.

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need," in Proceedings of the 2017 Advances in Neural Information Processing Systems (NeurIPS), pp. 5998–6008, 2017.

[19] S. Liu, D. Lin, and P. Luo, "Deep Learning-Based Medical Image Segmentation: A Survey," IEEE Transactions on Medical Imaging, vol. 42, no. 1, pp. 1–24, 2023.

[20] S. Razzak, S. Naz, and A. Zaib, "Deep Learning for Medical Image Processing: Overview, Challenges and Future," Neural Computing and Applications, vol. 32, no. 12, pp. 1–29, Springer, 2020.

[21] P. Lakhani and B. Sundaram, "Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis," Radiology: Artificial Intelligence, vol. 1, no. 2, pp. 1–10, 2019.

[22] J. Liu, Y. Pan, J. Li, and C. Tang, "A Multi-Scale CNN Model for Multi-Class Brain Tumor Classification," Frontiers in Neuroscience, vol. 14, pp. 1–10, 2020.

[23] H. Chen, C. Qi, J. Zhang, and Z. Wu, "Fusion of CNN and Transformer Features for Medical Image Classification," in Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, pp. 2432–2436, IEEE, 202

[24] J. Lin, Q. Song, X. Chen, and D. Xu, "Explainable Artificial Intelligence in Medical Imaging: A Review," Artificial Intelligence in Medicine, vol. 134, pp. 102–110, Elsevier, 2023.

[25] M. Schuld, I. Sinayskiy, and F. Petruccione, "The Quest for a Quantum Neural Network," Quantum Information Processing, vol. 13, no. 11, pp. 2567–2586, Springer, 2014.

[26] E. Henderson, R. Patel, and L. Xu, "Quantum-Inspired Optimization Techniques for Deep Learning Models," in Proceedings of the 2021 IEEE International Conference on Quantum Computing and Engineering (QCE), Broomfield, USA, pp. 152–159, IEEE, 2021.

[27] Z. Chen, T. Xu, and M. Wang, "Hybrid CNN–Transformer Network for COVID-19 Detection in Chest X-rays," Computers in Biology and Medicine, vol. 145, pp. 105–115, Elsevier, 2022.

[28] P. M. Reyes, K. Kim, and S. Lee, "Enhancing Medical Image Diagnosis Using Explainable Vision Transformers," in Proceedings of the 2023 IEEE International Symposium on Biomedical Imaging (ISBI), Cartagena, Colombia, pp. 504–509, IEEE, 2023.

[29] R. Singh, D. Sharma, and P. Kaur, "Quantum Deep Learning for Medical Image Analysis: A Review," IEEE Access, vol. 11, pp. 35421–35435, 2023.

[30] S. Kumar, N. Bhatia, and V. Sharma, "A Comparative Study of Deep CNN Architectures for Multi-Modal Disease Detection," in Proceedings of the 2024 International Conference on Intelligent Computing and Data Analytics (ICICDA), Singapore, pp. 87–93, Springer, 2024.