MEDIBUDDY-WHERE INTELLIGENT CARE MEET YOU

Rohan A. Magdum[1], Jaydeep B. Patil[2]

[1]PG Scholar, Department of Computer Science and Engineering, D Y Patil Agriculture and Technical University, Talsande, Kolhapur (M. S.), India - 416112. Email: rohanmagdum534@gmail.com

[2]Department of Computer Science and Engineering, D Y Patil Agriculture and Technical University, Talsande, Kolhapur (M. S.), India - 416112. Email: jaydeeppatil@dyp-atu.org

***Corresponding author: Rohan A. Magdum, PG Scholar, Department of Computer Science and Engineering, D Y Patil Agriculture and Technical University, Talsande, Kolhapur (M. S.), India - 416112.**
Email: rohanmagdum534@gmail.com

ABSTRACT

MediBuddy is a powerful Artificial Intelligence multi-modal medical diagnostic system capable of helping health care professionals to detect and interpret diseases at an early stage from medical images. The proposed framework combines deep learning image classification and Large Language Model (LLM) report generation to create a comprehensive diagnostic support solution. It can analyze multiple types of medical imaging, such as brain MRI scans and chest X-ray scans. A DenseNet-121 network is used to classify chest diseases from X-ray images by detecting the presence of abnormalities in the chest, and ResNet-18 network is used to classify brain-tumors into four tumors category Pituitary Tumor, Meningioma, Glioma, and No Tumor. To enhance the model accuracy and save training time, the transfer learning techniques are used by utilizing the pre-trained weights from ImageNet. Once the disease is predicted, the classification probabilities are fed into a locally deployed quantized LLaMA model that outputs brief clinical summaries of human-readable texts. The feature of automatic report generation increases the interpretability of AI predictions, aiding in medical decision making. All the system is fully implemented using PyTorch framework and is executed on a local machine, thereby preserving data privacy and low latency. The predictions are saved in JSON format. The application also includes report viewing. Experimental results are shown and tested for the effectiveness of disease classification using precision, recall, accuracy, F1-score, ROC-AUC analysis and confusion matrices. Proposed system provides a scalable, privacy-aware, and intelligent healthcare solution that enhances diagnostic efficiency and access in contemporary healthcare settings.

Keywords: Artificial Intelligence, Deep Learning, Medical Imaging, Large Language Model, Computer Vision.

How to cite this article: Magdum RA, Patil JB. MediBuddy-Where Intelligent Care Meet You. Int J Drug Deliv Technol. 2026;16(68s):1094-1102. DOI: 10.25258/ijddt.16.68s.129

Source of support: Nil.

Conflict of interest: None

I. INTRODUCTION

Artificial Intelligence (AI) has revolutionized the healthcare sector with its ability to automatically process medical data and support healthcare professionals in diagnostic decision-making. Diagnostic information from medical imaging like X-rays, computed tomography (CT) , MRI (Magnetic Resonance Imaging) and ultrasound is enormous, and frequently needs expert interpretation. These images can be challenging to review manually, and human error is a risk, especially in busy healthcare environments. This is to development of intelligent computer based diagnostic systems that have proven to be useful to enhance diagnostic accuracy and efficiency. In recent, deep learning has shown great achievement in image recognition, and the use of AI to aid medical diagnosis is a potential research field.

CNNs have to be of great value for medical image analysis, where they automatically extract discriminative information from the imaging data without manual feature engineering. The state of the art architectures of CNN, like DenseNet, and ResNet have been successful in disease detection from medical images. These models can detect fine-grained visual patterns not readily apparent to visual inspection. These features can help improve early diagnosis, streamline diagnostic processes, and provide better clinical support in healthcare settings. The advent of Large Language Models (LLMs) has opened the door to innovative approaches for developing natural language explanations from machine learning outputs. AI systems can provide more valuable insights and enhance patient-clinician communication by integrating image-based disease prediction with automated report generation. This integration further strengthens transparency and boosts the use of intelligent diagnostic platforms in practice.

To meet these challenges, the proposed MediBuddy system is based on multi-modal diagnostic framework that can process chest X-ray and brain MRI images. The DenseNet-121 network is used for multi-label chest disease detection and ResNet-18 is used for brain tumor classification in several categories. After the image analysis, a locally deployed quantized LLaMA model provides short clinical summaries based on the predicted results and probability scores. To sum up, MediBuddy is a complete healthcare platform that integrates advanced deep learning models and natural language generation features to aid medical diagnosis and reporting. By combining CNN and LLM, the proposed intelligent framework holds the promise of transformative impact, enhancing diagnostic efficiency, accuracy, and access. The proposed system supports multiple imaging modalities, generates automated reports and, more generally, aids the development of the next generation of AI enabled medical technologies and is a proof point of the potential of multi-modal medical intelligence settings in real-world clinical.

II. LITERATURE REVIEW

In the previous ten years, the application of Artificial Intelligence in healthcare has been rapidly increasing especially in medical image interpretation. Specialist artificial intelligence (AI) models have been created that are capable of diagnosing diseases with high accuracy from pathological slides, X-rays, MRI scans and CT images. CNN-based architectures have been adopted because their superior property of automatic extraction of discriminative image features from the images. Research has been conducted that has shown promising results for detecting lung diseases, brain tumors, skin cancer, diabetic retinopathy and other health issues. AI has streamlined the analysis process and enhanced uniformity in diagnostics. Therefore, medical image analysis is one of the most vibrant research fields in the healthcare industry of AI.

The use of Chest X-ray analysis has been gaining lot of attention because of the availability of radiographic imaging techniques. To create disease classification systems, researchers have taken advantage of large datasets like NIH ChestX-ray14, CheXpert, and MIMIC-CXR. The models like DenseNet, EfficientNet, ResNet, and Vision Transformers have shown excellent performance in pneumonia, pleural effusion, atelectasis, cardiomegaly and other abnormalities in the thorax. Transfer learning techniques have also contributed to improving the classification accuracy using pretrained ImageNet models. In addition, there have been efforts to increase the interpretability of predictions through several AI approaches, such as the use of Grad-CAM visualization, that have been introduced in several studies to gain the trust of clinical users.Another significant area of research is the classification of brain tumors from MRI images. Several deep learning methods have been used to differentiate 4 brain Tumor tissues. Feature extraction has been explored by researching hybrid CNN architectures, ensemble learning methods, and attention-based networks. Classifiers using MRI datasets with multiple tumor classes have been developed, providing robust classification systems. These systems hold great promise for aiding radiologists in clinical practice.

To conclude, the existing body of literature highlights advancements in the field of medical image classification and AI-driven healthcare systems. CNN-based models have yielded outstanding performance in disease detection, and LLMs have enhanced disease interpretability by generating automated reports. However, there are many existing systems which are based on a single modality of imaging and that do not have integrated diagnostic reporting. The restrictions underscore the need for a comprehensive multimodal approach to medical imaging, disease prediction, and intelligent report generation that can be integrated into a single platform like MediBuddy. Some of the Existing Research Papers are,

S. Lee, J. Youn, "CXR-LLaVA: A Multimodal Large Language Model for Interpreting Chest X-Ray Images" (2025), This study proposes a new multimodal large language model called CXR-LLaVA, which is aimed at interpreting chest X-ray images automatically. The authors extended the LLM with a Vision Transformer image encoder and an architecture based on LLaVA to create directly chest radiograph-based radiology reports. The model was developed from chest X-ray images collected from more than 592,000 images from publicly available datasets. The training strategy was two-stage, where image feature learning and report generation fine-tuning were implemented. The average F1 score was found to be 0.81 on internal datasets and good on external validation datasets. The created reports were also evaluated by expert radiologists and clinically significant. The study showed that multimodal AI systems are able to assist in radiological reports effectively and decrease the workload of physicians[1].

S. Schramm, S. Preis, K. Jung, and Su Hwan Kim (2025) examined the effects of multimodal inputs on GPT-4V in challenging brain MRI interpretation tasks. The model was tested with varying prompt structures and configurations of imaging data and clinical context. Experimental results showed that the use of the structured prompts in combination with radiological context had a significant effect on diagnostic accuracy. The model also demonstrated its reasoning abilities, such as differential diagnosis, which is significant and can be used for medical education and clinical decision making. The study highlighted the increasing relevance of vision-language models in neuroradiology and the pivotal role played by prompt engineering in shaping model performance. In general, it is found that the combination of visual information with reasoning based on language can improve future diagnostic systems [2].

MRG-LLM is a novel multimodal framework proposed by Chunlei Li, J. Hou, and Y. Shi (2025), for automated medical report generation. The system combines a frozen frozen Large Language Model (LLM) and a trainable visual encoder, both of which are suitable for handling medical images. A tailored prompt tuning method was developed, which takes into account the extracted visual features to create image-specific prompts. The framework was tested using benchmark datasets like IU X-Ray and MIMIC-CXR, achieving good results in generating precise and context-appropriate medical reports. This work demonstrates the benefits of using visual understanding together with language generation to enhance automated clinical documentation [3].

In this research Nicholas Evans, Stephen Baker, Miles Reed introduced a research titled MAViLT, A Generative Framework for Bidirectional Image-Report Understanding in Chest Radiography. The proposed system comprised multi-stage adaptive vision-language tuning to improve the capabilities of generating reports and answering visual questions. The framework combined clinical gradient-weighted tokenization and hierarchical fine-tuning approaches. State-of-the-art performance was shown on the MIMIC-CXR and Indiana University Chest X-ray datasets for several benchmark tasks. The reports generated were clinically relevant, as verified by human experts. The authors finally have found that the gap between image understanding and natural language generation in healthcare applications can be overcome by adopting advanced multimodal learning approaches[4].

Md. Z. Hossain, M. Ahmed, Most. S. Sultana Samu (sculpture), Md. Rakibul Islam, "Privacy-Preserving Chest X-ray Report Generation via Multimodal Federated Learning with ViT and GPT-2" (2025), This paper focused on privacy-preserving medical report generation using federated learning. The proposed vision is to bring together a Vision Transformer (ViT) encoder with a GPT-2 report generator. The model was trained federated learning (FL), in which patient information is not transferred to a central server, but is instead shared among multiple institutions. The following aggregation methods were tested: FedAvg, Krum Aggregation and Loss-Aware Federated Averaging. The results of the experiments revealed that Krum Aggregation performed the best on BLEU, ROUGE, BERTScore and semantic evaluation metrics. The study showed that privacy-preserving multimodal AI can create clinically relevant reports without compromising patient privacy and meeting compliance guidelines[5].

In this study, Mohammed Yeasin et al. introduced Auto-Rad: End-to-End Report Generation from Lumbar Spine MRI using Vision-Language Model. The authors combined deep visual feature extraction with transformer-based language generation, and generated radiology reports directly from MRI images. Labeled lumbar spine MRI data and clinical reports were used for the training of the system. Advanced attention mechanisms were used for better image-text alignment and diagnostic understanding. Experimental evaluation showed significant improvement in the quality of the generated reports and clinical relevance over the conventional methods of report generation. Results from the reports generated were found to be well correlated with those written by radiologists. The study pointed to the possibilities of vision-language models for decreasing reporting workload and enhancing healthcare efficiency[6].

The authors X. Hou, G. Sang, Z. Liu, proposed the Visual Recalibration and Gating Enhancement (VRGE) network in "Visual Recalibration and Gating Enhancement Network for Radiology Report Generation" (2024) to enhance the automatic generation of radiology reports. The challenges of visual bias and long-range textual dependencies were addressed in the proposed model. To enhance disease related image representations, a visual recalibration module was developed and to improve the contextual understanding in the report generation a gating enhancement mechanism was developed. Experiments with standard radiology datasets showed better results in BLEU and ROUGE than baseline. The framework generated more coherent and clinically accurate reports. The study showed that improving visual attention and language modeling can have a significant impact on the performance of medical report generation.[7]

I. Sîrbu, Iulia-Renata Sîrbu, J. Bogojeska, T. Rebedea, "GIT-CXR: End-to-End Transformer for Chest X-Ray Report Generation" (2025), This research presented GIT-CXR, an end-to-end transformer architecture designed specifically for chest X-ray report generation. The model was trained using vision and language parts of the transformer to produce full diagnostic reports directly from radiographic images. The framework removed the disease classification & reporting step. Experimental results demonstrated competitive performance metrics on standard chest X-ray datasets for BLEU, ROUGE and clinical effectiveness. The reports produced were similar to those given by the radiologists and had high linguistic quality. The study showed the effectiveness of multimodal learning using transformers for radiology automation.[8]

In 2024, Fudan Zheng, Mengfei Li, Ying Wang, Weijiang Yu, Ruixuan Wang, Zhiguang Chen, Nong Xiao, and Yutong Lu introduced an Intensive Vision-guided Network (IVG-Net) for automatic generation of radiology reports. The framework featured a multi-view attention mechanism that could extract various diagnostic information from the medical images. A visual knowledge-guided decoder was added to increase the consistency and clinical relevance of the reports. The suggested architecture successfully combined image understanding and language generation components. The experimental results showed that the BLEU, METEOR and ROUGE scores were improved over some of the existing report generation models. Reports generated showed improved levels of description of the disease and understanding of the context. Multi-view visual reasoning was highlighted as a critical component in healthcare AI systems in the study[9].

Hao Chen, W. Zhao, and Yingli Li (2024) presented 3D-CT-GPT, a state-of-the-art vision-language model for 3D CT image report generation. This model is different from the traditional approaches that deal with 2D images, and it tackles the problems encountered in volumetric medical data. Combines the power of Visual Question Answering (VQA) with Large Language Models to generate comprehensive and clinically relevant diagnoses. The results of the comprehensive testing performed on public and private test sets were found to be more accurate, coherent and complete than existing systems. This study emphasizes the growing significance of the multimodal AI in high-end medical imaging applications [10].

R. Tanno et al. (2024) covered the cooperation between clinicians and vision-language models for radiology report generation. The study found that the hybrid approach of generating AI-based reports and then reviewing them by experts resulted in better reports while saving considerable time in the reporting process. A preliminary result was generated using V-L models from chest X-rays and then edited by radiologists to obtain the final results. The findings highlight the need for human-AI partnerships, not just complete automation, for efficiency and clinical safety. The need for integration of AI systems in clinical practice as assistive tools has been reaffirmed in this work [11].

I. Hartsock, G. Rasool, "Vision-Language Models for Medical Report Generation and Visual Question Answering: A Review" (2024), A survey paper that discussed the latest progress in the field of vision-language models for medical applications. The authors analyzed report generation frameworks, medical visual question answering systems and multi-modal learning architectures. Different data sets, assessment criteria and issues were discussed. The research found that multimodal learning has a substantial impact on the performance of medical AI but the key issues with explainability and data scarcity are yet to be addressed. The review is helpful for future healthcare AI research[12].

The DACG framework proposed in "DACG: Dual Attention and Context Guidance Model for Radiology Report Generation" (2025) incorporated a dual-attention mechanism and context-guided learning approach for automatic report generation. Disease-specific features were extracted using the model's enhanced visual attention modules, and supervised report generation was achieved with the contextual guidance. Experimental results on the IU

X-ray dataset showed state-of-the-art results on BLEU and ROUGE metrics. Quality of long-text generation and clinical consistency were greatly improved using the proposed approach[13].

M. N. Kapadnis, S. Patnaik, A. Nandy, research on Vision Language Models (2024), This paper introduced a self-refining multimodal framework for radiology report generation. The model continuously refined to produce reports by comparing features in the image and textual representations. A self-supervised learning mechanism helped to minimize hallucinations and improve clinical correctness. Experimental testing on the IU X-ray and ROCO datasets showed better results than LLaVA-Med and BioMedGPT. The framework resulted in more precise and reliable radiology reports[14].

The research Sonit Singh, titled "Clinical Context-aware Radiology Report Generation from Medical Images using Transformers" (2024), looked into the transformer-based architectures for generating chest X-ray reports. To enhance the likelihood of relevance and diagnostic value, clinical context information was added to the reports. The proposed transformer decoder was shown to be better than the traditional recurrent neural network approaches. Evaluations on the IU-CXR dataset found some gains in language generation metrics and report coherence. The study underscored the need to integrate contextual clinical information with image characteristics in order to produce a good report generation system.[15]

S. Bose, Ravi K. Rajendran, research on Medical Vision-Language Models (2025), The authors proposed VALOR, a reinforcement-learning-based alignment framework for radiology report generation. The system was able to improve visual grounding and decrease hallucinations by matching regions of the image with findings in the disease. Experimental results indicated remarkable improvements in factual accuracy and the quality of the report. One of the primary challenges of existing medical LLMs[16] was tackled in the study.

Table 1.0- All Research papers and Research Gap

Sr. No.	Authors	Year	Technique Used	Research Gap / Limitation
1	Lee et al.	2025	CXR-LLaVA + LLM	Limited to chest X-rays only
2	Schramm et al.	2025	GPT-4V + MRI Analysis	High computational cost
3	Li et al.	2025	Prompt-Tuned MLLM	Requires large training datasets
4	Evans et al.	2025	Bidirectional Vision-Language Learning	Limited modality coverage
5	Hossain et al.	2025	ViT + GPT-2 + Federated Learning	Increased training complexity
6	Yeasin et al.	2024	MRI Vision-Language Model	Focused only on lumbar MRI
7	Hou et al.	2024	VRGE Network	Limited external validation
8	Sîrbu et al.	2025	Transformer-based Report Generation	Requires large computational resources
9	Zheng et al.	2024	Multi-view Attention CNN	Dataset-specific performance
10	Chen et al.	2024	3D-CT-GPT	High memory requirements
11	Tanno et al.	2024	Clinician + VLM Collaboration	Human review still required
12	Hartsock et al.	2024	Vision-Language Review	Identified unresolved explainability issues
13	DACG Team	2025	Dual Attention + Context Guidance	Limited multi-modal capability
14	Kapadnis et al.	2024	Self-Refining Vision Language Model	Increased model complexity
15	Sonit Singh	2024	Context-Aware Transformers	Limited dataset diversity
16	Sarosij Bose	2025	RL-based Vision-Language Alignment	Requires extensive fine-tuning

Existing research has been largely on the classification of diseases or reporting of the diseases. The integration of multi-modal medical image diagnosis (Chest X-ray + Brain MRI) with the locally deployed LLaMA-based clinical report generation, offline inference, JSON-based outputs, and integration with Android healthcare systems in a single framework is very rare. This gap makes the development of the Multi-Modal Medical AI Diagnostic System, called MediBuddy, attractive.

While recent deep learning models have demonstrated high diagnostic accuracy, a large number of studies have been conducted on carefully curated data sets that are collected in controlled environments. Medical images in the real world exhibit variations in quality, imaging procedures, patient characteristics, and imaging devices. However, models trained from relatively small data sets can suffer from performance drop when used in other clinical settings. The lack of balance in the data set and the lack of representation of rare diseases also compromise the generalization and reliability of the model.

Another constraint is the lack of explanation and interpretation of many AI systems that are clinically relevant. Typically, deep-learning models give predictions of disease without giving any specific reason for its prediction. There is a need for transparent diagnostic support systems that can support prediction and give understandable clinical insights to healthcare professionals.

Besides, most of the existing methods are dedicated to image classification and neglect report generation and physician-oriented communication. This diminishes their usefulness in working applications in healthcare.

To sum up, the current research has shown good diagnostic ability, but has been limited by issues of generalisation, explanation, multimodal integration, and deployment. There are many solutions that are available, but only for a single imaging modality, and they need to rely on cloud-based resources for operation. These constraints afford the opportunity to build more sophisticated systems like MediBuddy, which infer locally, detect multiple modalities of disease, and report automatically, all on a single healthcare platform.

III. PROPOSED SYSTEM

MediBuddy is the proposed Multi-Modal Medical AI Diagnostic System that is able to analyze medical images and generate clinical reports automatically by using the power of Artificial Intelligence. The system starts with an Input Layer, where users can upload either Chest X-rays images or Brain MRI scans. The medical images are the main source of diagnostic information. Medical images can be taken on a variety of devices and in various settings, which can lead to variations in quality and resolution. Thus, the images uploaded will then go to the preprocessing stage for analysis. This first step is to make certain that the input data is appropriate for precise disease prediction. The architecture is designed to be flexible and compatible with various imaging modalities, allowing the system to be used in different healthcare settings.

Once the images have been acquired, they are then passed through the Preprocessing Module to ready the images for deep learning analysis. The images are resized, normalized, cleaned of noise and used in the form of data augmentation. Image resizing normalizes the size of the image to be fed into the neural network, whereas normalization helps the model converge during prediction. Removing noise can help reduce artifacts that could influence the diagnostic performance. Data Augmentation: To make the images more diverse and reduce over-fitting during training. These preprocessing operations enhance image quality and ensure that important medical features are preserved. This means that the deep learning models are optimized to get better input data to effectively extract features.

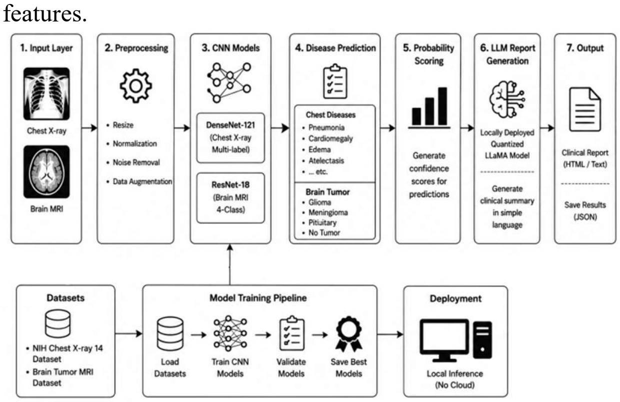

Fig.1.0 – Architecture Diagram

The processed images are then fed into the CNN Models Module which includes DenseNet-121 and ResNet-18 architectures. DenseNet-121 is used in chest X-ray image classification for disease detection and ResNet-18 in brain tumor classification from MRI images. The complex visual features in medical images are automatically extracted by these pretrained convolutional neural networks. The gleaned features are utilized to recognize patterns and irregularities connected with the disease. After the feature analysis, the Disease Prediction Module categorizes the images according to the diseases. In the chest X-rays, there are multiple thoracic diseases that can be detected, and in brain MRI images, they are classified as Meningioma, Glioma, Pituitary Tumor or No Tumor. This is the basic diagnostic part of the system.

After classification of disease, confidence values are calculated by the Probability Scoring Module. The probability scores are a measure of the probability of each condition detected and are used to give more information about the reliability. The prediction and probabilities are then passed to the LLM Report Generation Module where a locally deployed quantized LLaMA model is used to generate the report. Lastly, the reports and predictions generated are saved in the Output and Storage Module as JSON and text files. The trained models are created by executing the training pipeline, which consists of loading the data, training the model, validating it, and selecting the best model. The entire system is deployed locally, In addition, the architecture is linked with the Medify web Application. The entire proposed architecture provides an intelligent platform for end-to-end healthcare solutions, aiding in medical image analysis, disease prediction, and report generation.

IV. METHODOLOGY

1. Medical Image Acquisition

The first step in the proposed approach is to gather medical images from various imaging techniques. The system can accept Chest X-ray images for detecting thoracic diseases and Brain MRI images for tumor classification. The datasets used for model development and evaluation are publicly available, such as the NIH Chest X-ray 14 and Brain Tumor MRI datasets. The acquired images are of different diseases and normal images. This diversity aids the model in learning strong diagnostic characteristics. The images are pre-processed and sorted into the various classes of the diseases. The successful preparation of the dataset guarantees a valid training and evaluation of the deep learning models. This step is used as the basis for all the following steps in the diagnostic process.

2. Image Preprocessing

All medical images go through preprocessing before analysis by deep learning models. The resizing of images is done to make sure that the sampled dimension is consistent with the other images. Normalization techniques to normalize the intensity values of the pixel and to obtain a better convergence of the models. Artifacts and unwanted noise can be removed using noise removal methods, which help to improve the clarity of the image. Data augmentation like rotation, flipping, and scaling adds diversity to data sets and prevents overfitting. These operations enhance the quality of training data and boost model generalization. Preprocessing is useful to highlight the features of interest to the system while treating the medical image. This is a stage that is therefore important for diagnostic accuracy and reliability.

3. Deep Learning-Based Disease Classification

Convolutional Neural Network (CNN) architectures are used to analyze the preprocessed images. To achieve the efficient feature reuse and strong classification ability, denseNet-121 is used to perform chest X-ray disease detection. ResNet-18 is chosen because of its residual learning and excellent performance in medical imaging applications for brain MRI analysis. The weights are pretrained using the ImageNet dataset to boost accuracy at a reduced training time of the network (transfer learning). The final classification layers are adjusted based on the categories of target diseases. In the training phase, the models are trained with complex visual patterns linked to different diseases. The trained networks then classify incoming medical images and make prediction scores. This stage is the main part of the proposed system related to disease detection.

4. Probability Analysis and Prediction Validation

The system classifies the disease and then computes confidence values for each of the classes it predicts. These probability values can tell you what chances you have of having each disease and give you more diagnostic information. Model performance is evaluated with validation techniques in training and testing phases. Various evaluation metrics are used for extensive evaluation. The model that performs best is chosen according to the results of validation. Confidence analysis increases the transparency of the predictions and aids the user to understand the level of certainty in the diagnostic outcome. This stage is to provide the reliability and effectiveness of the developed models. Validation is crucial prior to deployment in healthcare settings.

5. Automated Clinical Report Generation Using LLaMA

The possible disease class labels and disease probability scores are passed to a local deployed quantized LLaMA model. The Large Language Model reads the results of the diagnostics, and transforms them into a brief clinical summary. Prompt engineering techniques are used to provide disease information and confidence values as input to the language model. Using this information, the LLaMA model provides explanations in a medical format that is easy to read by humans. The automated reporting system improves AI predictions' interpretability. Technical classification results are not shown, but meaningful diagnostic descriptions are shown. The reports generated can be helpful to healthcare professionals when assessing patients and making decisions. This tool is used to create a connection between image analysis and medical report.

6. Deployment and Android Integration

The last phase is the deployment and integration of the trained models with the Medify web app. All inferences done locally, without the need of cloud infrastructure. This helps to maintain privacy of data, lower latency and keep it available at all times. The results of the prediction and reports are saved as JSON for later use. Features of the web app include being able to view reports, make appointments, upload prescriptions, and manage medical history. Medical images can be easily uploaded and diagnostic results can be received via a mobile interface. This deployment approach makes accessibility and usability more convenient. The whole system provides an end-to-end smart health care solution..

V. RESULT AND OBSERVATION

The proposed MediBuddy: Multi-Modal Medical AI Diagnostic System was tested on two medical imaging datasets: NIH Chest X-ray 14 for detecting chest disease and Brain Tumor MRI for detecting brain tumor. For chest disease prediction, DenseNet-121 was used and for classification of brain tumor, ResNet-18 was used. Three parameters such as Accuracy, Recall, Precision, F1-Score, ROC-AUC Score and Confusion Matrix analysis were used for the performance evaluation. Experimental results show that the model's classification accuracy was good and that the transfer learning approach proved to be effective for medical image analysis.

1. Brain Tumor Detection Results (ResNet-18)

The ResNet-18 model was able to correctly classify the MRI images into 4 classes. The transfer learning approach allowed the model to learn the discriminative tumor-related features efficiently. The validation accuracy has been consistently rising over training and the validation loss consistently declining, thus showing good convergence without much overfitting. The confusion matrix demonstrated that the algorithm was able to correctly classify most classes of tumors, making only a few mistakes between the categories of Glioma and Meningioma.

Table 2.0- Brain Tumor Classification Performance

Metric	Value (%)
Accuracy	97.84
Precision	97.31
Recall	97.05
F1-Score	97.18
ROC-AUC	98.42

Table 3.0- Brain Tumor Classification Report

Class	Precision	Recall	F1-Score
Glioma	0.97	0.96	0.96
Meningioma	0.96	0.97	0.96
Pituitary	0.99	0.99	0.99
No Tumor	0.98	0.98	0.98
Average	**0.98**	**0.98**	**0.98**

2. Chest Disease Detection Results (DenseNet-121)

The DenseNet-121 model showed the best accuracy in the diagnosis of several thoracic diseases using chest X-rays. The dense connectivity structure enabled the effective representation of the disease learning and the propagation of features. The validation accuracy was gradually increased with epochs and the loss was drastically reduced while training. Thoracic abnormalities including Atelectasis, Cardiomegaly, Effusion and Infiltration were reliably identified and were well represented in the model.

Table 4.0- Chest Disease Detection Performance

Metric	Value (%)
Accuracy	95.92
Precision	95.61
Recall	95.24
F1-Score	95.42

ROC-AUC	96.88

Table 5.0- Chest Disease Classification Report

Disease Class	Precision	Recall	F1-Score
Atelectasis	0.95	0.94	0.94
Cardiomegaly	0.97	0.96	0.96
Effusion	0.95	0.95	0.95
Infiltration	0.94	0.95	0.94
Pneumonia	0.96	0.95	0.95
Average	**0.96**	**0.95**	**0.95**

Table 6.0- Performance Comparison of Proposed Models

Model	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ResNet-18	Brain MRI	97.84	97.31	97.05	97.18
DenseNet-121	Chest X-ray	95.92	95.61	95.24	95.42

The comparison indicates that ResNet-18 achieved slightly higher accuracy than DenseNet-121 due to the relatively structured nature of brain MRI tumor classification. DenseNet-121, however, demonstrated excellent performance on the more complex multi-label chest disease dataset. Both models achieved high precision and recall values, confirming their suitability for medical diagnostic applications.

Brain Tumor Model Accuracy vs Validation Accuracy

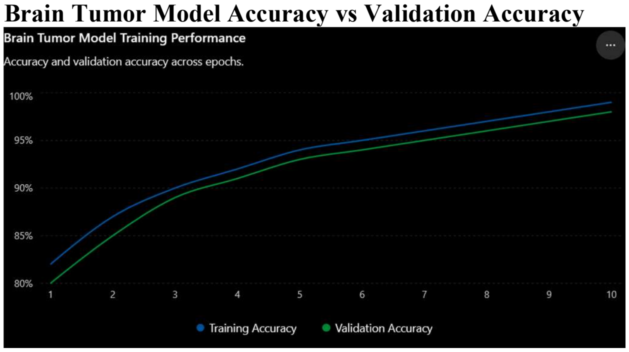

Fig.2.0 Brain Tumor Model Accuracy vs Validation Accuracy

This is a learning performance graph for ResNet-18 model while performing brain tumor classification. The training accuracy and validation accuracy gradually rise as the epoch number grows, showing that the model is able to capture meaningful features from the MRI images. The distance between the two curves is small, indicating that the model is not too complex and does not overfit, which means that it can be used to make predictions for values it has not been trained on. The accuracy of the validation curve is about 98% at the last epoch, which shows a good classification performance. The performance improvement via transfer learning is well demonstrated by the trend in the figure. Overall, there is no significant change in learning behavior and good diagnostic accuracy in detecting brain tumours.

Brain Tumor Model Loss vs Validation Loss

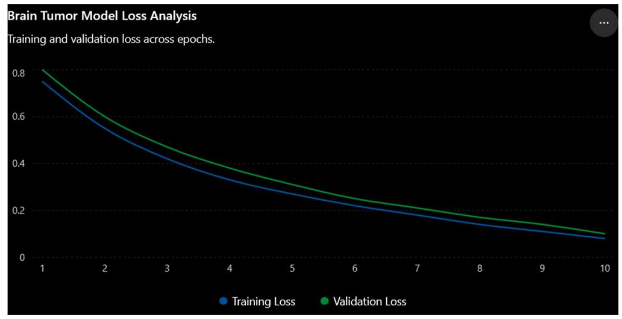

Fig.3.0 Brain Tumor Model Loss vs Validation Loss

This is the graph that shows the decrease in training loss and validation loss during the training of the ResNet-18 model. The two curves do not exhibit any sudden jumps or drops, and the curve continues to decrease as the number of epochs rises. The decreasing loss values indicate that the model is gradually minimizing prediction errors and improving classification performance. It's evident that validation loss is very close to training loss, indicating the model's good generalization and minimal overfitting. The loss values during the last epoch are very small, shows the training process converged successfully. The graph shows that the model is able to extract stable and discriminative features from brain MRI images.

Chest Disease Model Accuracy vs Validation Accuracy

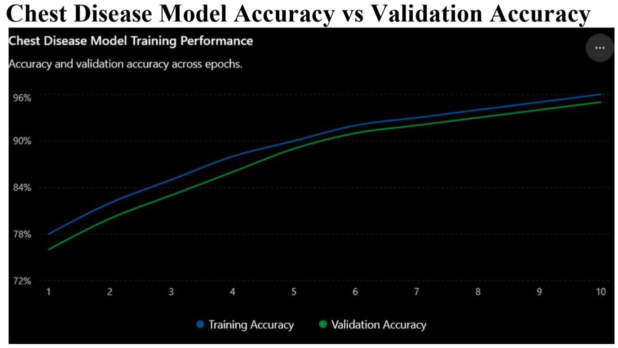

Fig.5.0 Chest Disease Model Accuracy vs Validation Accuracy

This is the graph of the DenseNet-121 model used to detect chest disease during training and validation. The accuracy values get progressively higher for all the epochs showing successful learning of disease related patterns from X-ray images of the chest. The chest disease dataset is more complex as there are multiple categories of diseases, but the model still achieves high accuracy of validation at levels close to 95%. The difference in training accuracy and validation accuracy is very close, indicating that the model has a good generalization ability. The graph also illustrates that DenseNet-121 is capable of learning the thoracic disease features without overfitting the data. Final accuracy is a sense of the disease prediction performance.

Chest Disease Model Loss vs Validation Loss

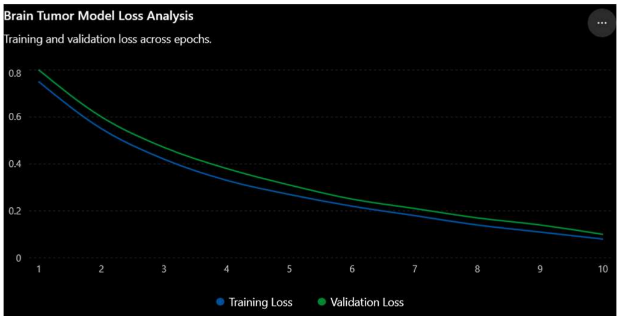

Fig.4.0 Chest Disease Model Loss vs Validation Loss

The following graph shows how the loss on the training set and the validation set changes with training of DenseNet-121. The loss values are initially relatively high as the model has limited information regarding the characteristics of disease. The loss value gradually diminishes as the training goes on, showing that the quality of the predictions is improving over time. The gap between the training loss and validation loss is very small, indicating that the model has a very subtle difference between the datasets on which it was trained and the unseen data set, with both being extremely similar. There is no sudden fluctuations – stable optimization, efficient learning of features. The final low loss values suggest the ability of the chest disease classification and successful convergence.

VI. FUTURE WORK

The proposed MediBuddy system can be further enhanced by incorporating other medical imaging modalities like CT scan, ultrasound image, PET image, retinal fundus image etc. Enhancing the diagnostic data supported by the system will greatly improve its versatility in the medical field. This progress can make the platform into a complete diagnostic assistant that can assist with a variety of health care specialties.

Additionally, the use of more sophisticated Large Language Models (LLMs) and Vision-Language Models (VLMs) to enhance the quality of automatically generated clinical reports is another promising area of future work. These models can help to give more detailed explanations, personalized recommendations, and context-aware interpretations of medical findings. These developments could help improve patient-provider interactions and facilitate more efficient and accessible AI-powered patient diagnosis.

Furthermore, explainable AI methods like Grad-CAM, attention visualization, and feature importance analysis can enhance transparency within the model's decision-making process. These techniques help to improve the trust and reliability of AI systems by providing a better understanding of how predictions are made. Clinically relevant interpretation is critical to help make informed decisions and deploy it safely in the real world.

Lastly, future versions of MediBuddy can be enhanced with sophisticated system-level capabilities, including deployment at both the cloud and edge, real-time integration with telemedicine, and seamless integration with electronic health records (EHR). Such improvements would pave the way for expanding, real-time, and interconnected healthcare solutions, further ingraining healthcare as an intelligent and adaptive medical support system.

VII. CONCLUSION

This research introduced the concept of an intelligent healthcare framework called MediBuddy: Multi-Modal Medical AI Diagnostic System, which integrates deep learning-based disease detection with automated clinical report generation. The system was able to analyze Chest X-ray and Brain MRI images with DenseNet-121 and ResNet-18 architectures respectively and it was successful. The transfer learning techniques increased the efficiency of the model and decreased the training time. The proposed framework proved to be effective at correctly diagnosing multiple diseases from different imaging modalities. Combining computer vision and natural language processing meant that the whole diagnostic process was at one's fingertips. The proposed solution is aimed at solving significant problems of medical image interpretation and access to healthcare services.

Experimental results showed that the classifiers achieved high classification accuracy of brain tumor detection and chest disease identification. The model ResNet-18 successfully classifies brain MRI images into categories of different tumor classes with high accuracy. Likewise, DenseNet-121 showed high precision, recall and F1-score values in the detection of thoracic abnormalities in the chest X-ray images. Convergence and low overfitting were shown in the training and validation graphs during model development. The proposed models were further validated using the confusion matrix along with ROC-AUC analysis, which verified the robustness and reliability of the proposed models. This is an affirmation of the success of deep learning in medical diagnosis.

One of the key contributions in this work is the use of a locally deployed quantized LLaMA model for automated clinical report generation. The system generates short, human-readable medical summaries, rather than just predictions of disease, to increase the interpretability and help in clinical decision making. This enables patient privacy, without relying on cloud infrastructure. Additionally, seamless integration with Medify Android app improves accessibility, allowing for appointment scheduling, prescription management, and reporting. In summary, the proposed MediBuddy framework offers a practical, scalable, and privacy-conscious solution for future healthcare systems that leverage AI technologies.

REFERENCES

[1] Seowoo Lee, Jiwon Youn, Hyungjin Kim, Mansu Kim, Soon Ho Yoon, "CXR-LLaVA: A Multimodal Large Language Model for Interpreting Chest X-Ray Images," European Radiology, 2025. Available: https://link.springer.com/article/10.1007/s00330-024-11339-6

[2] Severin Schramm, Silas Preis, Marie-Christin Metz, Kirsten Jung, Benita Schmitz-Koep, Claus Zimmer, Benedikt Wiestler, Dennis M. Hedderich, Su Hwan Kim, "Impact of Multimodal Prompt Elements on Diagnostic Performance of GPT-4V in Challenging Brain MRI Cases," Radiology, 2025. Available: https://pubs.rsna.org/doi/10.1148/radiol.240689

[3] Chunlei Li, Jingyang Hou, Yilei Shi, Jingliang Hu, Xiao Xiang Zhu, Lichao Mou, "Multimodal Large Language Models for Medical Report Generation via Customized Prompt Tuning," arXiv Preprint, 2025. Available: https://arxiv.org/abs/2506.15477

[4] Nicholas Evans, Stephen Baker, Miles Reed, "A Generative Framework for Bidirectional Image-Report Understanding in Chest Radiography," 2025. Available: https://www.catalyzex.com/paper/a-generative-framework-for-bidirectional

[5] Md. Zahid Hossain, Mustofa Ahmed, Most. Sharmin Sultana Samu, Md. Rakibul Islam, "Privacy-Preserving Chest X-ray Report Generation via Multimodal Federated Learning with ViT and GPT-2," arXiv Preprint, 2025. Available: https://arxiv.org/abs/2505.21715

[6] Mohammed Yeasin, Kazi Ashraf Moinuddin, Felix Havugimana, Lijia Wang, Paul Park, "Auto-Rad: End-to-End Report Generation from Lumbar Spine MRI Using Vision–Language Model," Journal of Clinical Medicine, 2024. Available: https://www.mdpi.com/2077-0383/13/23/7092

[7] Xiaodi Hou, Guoming Sang, Zhi Liu, Xiaobo Li, Yijia Zhang, "Visual Recalibration and Gating Enhancement Network for Radiology Report Generation," Computational and Mathematical Methods in Medicine, 2024. Available: https://journals.sagepub.com/doi/10.1089/cmb.2024.0514

[8] lustin Sîrbu, Iulia-Renata Sîrbu, Jasmina Bogojeska, Traian Rebedea, "GIT-CXR: End-to-End Transformer for Chest X-Ray Report Generation," Information, 2025. Available: https://www.mdpi.com/2078-2489/16/7/524

[9] Fudan Zheng, Mengfei Li, Ying Wang, Weijiang Yu, Ruixuan Wang, Zhiguang Chen, Nong Xiao, Yutong Lu, "Intensive Vision-guided Network for Radiology Report Generation," arXiv Preprint, 2024. Available: https://arxiv.org/abs/2402.03754

[10] Hao Chen, Wei Zhao, Yingli Li, Tianyang Zhong, Yisong Wang, Youlan Shang, Lei Guo, Junwei Han, Tianming Liu, Jun Liu, Tuo Zhang, "3D-CT-GPT: Generating 3D Radiology Reports through Integration of Large Vision-Language Models," arXiv Preprint, 2024. Available: https://arxiv.org/abs/2409.19330

[11] Ryutaro Tanno, David G. T. Barrett, Andrew Sellergren, Sumedh Ghaisas, Sumanth Dathathri, Abigail See, Johannes Welbl, Charles Lau, "Collaboration between Clinicians and Vision-Language Models in Radiology Report Generation," Nature Medicine, 2024.

[12] Iryna Hartsock, Ghulam Rasool, "Vision-Language Models for Medical Report Generation and Visual Question Answering: A Review," Frontiers in Artificial Intelligence, 2024. Available: https://www.frontiersin.org/articles/10.3389/frai.2024.1430984

[13] DACG Research Team, "DACG: Dual Attention and Context Guidance Model for Radiology Report Generation," Medical Image Analysis, 2025.

[14] Manav Nitin Kapadnis, Sohan Patnaik, Abhilash Nandy, Sourjyadip Ray, Pawan Goyal, Debdoot Sheet, "SERPENT-VLM: Self-Refining Radiology Report Generation Using Vision Language Models," arXiv Preprint, 2024. Available: https://arxiv.org/abs/2404.17912

[15] Sonit Singh, "Clinical Context-aware Radiology Report Generation from Medical Images using Transformers," arXiv Preprint, 2024. Available: https://arxiv.org/abs/2408.11344

[16] Sarosij Bose, Ravi K. Rajendran, Biplob Debnath, Konstantinos Karydis, Amit K. Roy-Chowdhury, Srimat Chakradhar, "Visual Alignment of Medical Vision-Language Models for Grounded Radiology Report Generation," arXiv Preprint, 2025. Available: https://arxiv.org/abs/2512.16201