MULTI-MODAL MEDICAL IMAGE ANALYSIS SYSTEM FOR AUTOMATED DISEASE DETECTION USING DEEP LEARNING

A PROJECT REPORT

Submitted By

AVINASH S	**2116220701034**
BHARATH KUMAR L	**2116220701043**
CHANDDRAPRAKASH S	**2116220701049**

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

THANDALAM : CHENNAI 602 105

MAY 2025

RAJALAKSHMI ENGINEERING COLLEGE, CHENNAI

BONAFIDE CERTIFICATE

Certified that this Project titled **" MULTI-MODAL MEDICAL IMAGE ANALYSIS SYSTEM FOR AUTOMATED DISEASE DETECTION USING DEEP LEARNING"** is the bonafide work of **"AVINASH S (2116220701034), BHARATH KUMAR L(2116220701043) CHANDDRAPRAKASH S (2116220701049)"** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE	**SIGNATURE**
Dr. P. Kumar., M.E., Ph.D.,	Dr. M . Ayyadurai., M.E., Ph.D.,
HEAD OF THE DEPARTMENT	**SUPERVISOR**
Professor	Associate Professor
Department of Computer Science	Department of Computer Science
and Engineering,	and Engineering,
Rajalakshmi Engineering College,	Rajalakshmi Engineering College,
Chennai - 602 105.	Chennai - 602 105.

Submitted to Mini Project Viva-Voce Examination held on _______________

Internal Examiner	**External Examiner**

ABSTRACT

This project introduces a deep learning-based multi-modal medical image analysis system for automated disease detection using CT scans. It supports both image and video inputs and targets four medical conditions: lung pneumonia, brain strokes, kidney stones, and spine fractures. The backend is built using FastAPI, ensuring robust RESTful API services, while the frontend uses Next.js to provide a responsive, user-friendly interface. The system integrates a dynamic model loader and a video frame extraction pipeline, offering real-time analysis. It employs transfer learning with EfficientNetB0 as the backbone, along with custom DepthwiseConv2D layers for optimized medical image compatibility. A comprehensive preprocessing pipeline handles RGB conversion, resizing, normalization, and batch processing. Each model is tuned with custom loss functions and confidence thresholds. Predictions include class labels, confidence scores, and probability distributions. Video inputs are analyzed using a batch-wise frame inference mechanism with results aggregation. The system demonstrates high accuracy across all conditions, with image processing under one second and video processing at 2–3 seconds per frame. Its modular architecture allows for easy integration of additional diseases. The platform ensures secure uploads, confidence-based results, and visual feedback. Clinical relevance includes early disease detection, workflow integration, and decision support. The project sets a strong foundation for real-time, scalable, AI-assisted medical diagnostics in healthcare.

ACKNOWLEDGMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavor to put forth this report. Our sincere thanks to our Chairman **Mr. S. MEGANATHAN, B.E, F.I.E.**, our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN**, **Ph.D.,** for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P. KUMAR, M.E., Ph.D.**, Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guides **Dr. M. AYYADURAI** and **Dr. G.M. SASIKALA.** We are very glad to thank our Project Coordinator, **Dr. M. AYYADURAI** Associate Professor Department of Computer Science and Engineering for his useful tips during our review to build our project.

AVINASH S 2116220701034

BHARATH KUMAR L 2116220701043

CHANDDRAPRAKASH S 2116220701049

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

S.NO	ABBR	EXPANSION
1	CT	Computed Tomography
2	CNN	Convolutional Neural Network
3	API	Application Programming Interface
4	FastAPI	Fast Asynchronous Web Framework
5	GPU	Graphics Processing Unit
6	AUC	Area Under Curve
7	F1-score	Harmonic Mean of Precision and Recall
8	RGB	Red Green Blue
9	DICOM	Digital Imaging and Communications in Medicine
10	HOG	Histogram of Oriented Gradients
11	K-Fold	K-Fold Cross Validation
12	HTML	HyperText Markup Language
13	CSS	Cascading Style Sheets
14	REST	Representational State Transfer

CHAPTER 1

INTRODUCTION

1.1 GENERAL

Medical imaging is an essential part of modern diagnostics, which plays an important role in detecting and managing various health conditions. With the increasing availability of CT scan data, there is a growing demand for intelligent systems that can aid in rapid and accurate analysis. Traditional manual interpretation of medical images can be time consuming and may be prone to variability between radiologists. To resolve these challenges, the project proposes a multi-model medical image analysis system that is operated by deep learning.

The system is capable of analyzing both stable images and video scans to detect several diseases including lung pneumonia, brain stroke, kidney stones and spine fractures. It takes advantage of the Convisional Neural Network (CNN) with transfer learning to achieve high clinical accuracy. IFFICENETB0 is used as a backbone model, which is combined with custom layers for better compatibility with medical data.

On the backnd, the system is applied using Fastapi, which offers a scalable and fast restful API. CRONTEND is designed using next.JS, provides a comfortable user interface with real-time response and drag-and-drag uploads. Advanced features include video frame extraction, confidence-based prediction and batch processing. The modular design allows for integration of additional conditions in the future. The project displays the practical use of AI in healthcare, aimed at supporting physicians with accurate, sharp and reliable clinical insight.

1.2 OBJECTIVE

The purpose of this research is to create an AI-based multi-modal medical image analysis system capable of detecting several diseases from CT scan images and videos. It targets conditions such as lung pneumonia, brain stroke, kidney stones and spinal fractures using deep learning techniques. By taking advantage of transfer learning and custom CNN architecture, the system aims to obtain high classification accuracy with limited training data. A integrated pipeline efficiently handles both stable and dynamic inputs, including adaptive frame extraction and preprocessing. The backend has been applied with FastAPI for scalable and safe API services. Front &, developed using Next.JS, provides an responsible and user friendly interface. A confidence scoring mechanism with threshold-based predictions increases clinical reliability. The system to support additional medical conditions in the future is modular, real -time and expandable.

1.3 EXISTING SYSTEM

Current medical image analysis systems mainly focus on single-dissemination detection and rely too much on large annotated datasets. Most devices only analyze stable images, which lacks support for video input or real -time processing. Traditional systems often use traditional image processing or basic machine learning techniques, which limit clinical accuracy and flexibility. Many lack modular architecture, making it difficult to add support to new conditions. Back-end implementation is usually unbroken with limited scalability or API integration. User interfaces are often non-co-ordinated and do not support dynamic interaction or detailed results visualization. The confidence scoring is either absent or not clinically reliable. Overall, the existing systems are less adaptive, less efficient, and are not designed for multi-seed, multi-modal diagnostic functions.

CHAPTER 2

LITERATURE SURVEY

[1] The paper by Singh and Luxmi (2023) examines deep learning methods to automate blood cancer diagnosis along with detection tasks. The research interpretation of blood cell data works to improve medical diagnosis alongside clinical decision making. The implementation of convolutional neural networks enables medical professionals to detect malignant patterns in blood testing samples which benefits patient diagnosis. The study recognizes two main obstacles which include insufficient available data and the requirement of uniform datasets for achieving model stability across multiple population groups.

[2] The authors of Ejiyi et al. (2025) developed ResoMergeNet as a deep learning framework which performs multi-modality medical image classification specifically for detecting cataract alongside lung cancer and breast cancer. The model achieves better diagnosis precision and complete clinical evaluations by merging features between different imaging techniques. This patient-focused methodology holds excellent promise in developing diagnostic operational standards. The model works effectively only when input data from various imaging sources possesses consistent quality.

[3] The researchers Hamsagayathri and Vigneshwaran (2021) developed a machine learning system which uses patient-reported symptoms to forecast diseases. Through support vector machines and decision trees algorithms the model provides an invasive disease detection system which supports early-stage disease identification. The proposed method helps improve speed and efficiency when conducting initial assessments within medical clinics. The precision of the system depends entirely on patient-found symptom information which must be precise and comprehensive.

[4] The research by Archana, Govindraj, and Adhil (2024) demonstrates an improved strategy for CT scan stone annotation in kidney diagnosis. The researchers use innovative imaging approaches with annotation tools to enhance kidney stone detection processes together with diagnosis classification methods. Receiving diagnoses along with

customized therapy plans becomes more possible through this method. The system depends on superior imaging data quality while also facing variability issues when it comes to annotation standards.

[5] The researchers from Khan et al. (2025) establish an optimized deep learning model that performs complete medical image analysis between different modalities. Neural network architectures in the model allow its analysis of different medical images to improve diagnostic capabilities. The standardized method enables workflow optimization in hospital diagnostic settings. The model has limitations regarding its training efficiency due to the scarce availability of extensive annotated datasets spanning all medical imaging modalities.

[6] Dehbozorgi, Ryabchykov and Bocklitz (2025) performed a research comparative study which examined the features produced by statistical, radiomics and deep learning methods for medical image classification. Each method receives assessment based on their performance capabilities within radiological and optical modalities for providing insights regarding their unique benefits. This evaluation helps clinical personnel choose the most suitable feature extraction methods needed for particular medical applications. A standardized evaluation metric system is needed to enable proper comparison of methods according to the findings of this study.

[7] The authors Richa and Patro (2025) developed a deep learning framework that seeks to enhance breast cancer early detection through medical imaging. A diagnostic precision improvement happens through the combination of convolutional neural networks with feature selection methods in the proposed model. Such an approach has the ability to enable quick intervention and enhance patient outcomes. The model shows two primary weaknesses because it requires top-notch imaging input combined with extensive testing in numerous population groups.

[8] Rajasekar et al. (2023) perform lung cancer disease prediction by processing CT scans and histopathological images with deep learning algorithms. Multiple neural network structures assist medical feature extraction and analysis in this study to enhance diagnosis

capability. The combination of multiple medical approaches through this methodology proves effective at discovering lung cancer at more advanced stages of development. Performance of this model gets affected by the diversity in imaging data as well as the requirement of extensive training datasets.

[9] Rajput and Subasi (2023) dedicated their research to detect lung cancer through histopathological lung tissue images with the help of deep learning. This chapter describes how convolutional neural networks lead to improved diagnosis accuracy through their identification of tissue sample malignant patterns. This method supports pathologists in clinical decision-making processes. The approach has limitations because it needs high-resolution images together with potentially diverse methods of tissue sample preparation.

[10] Albuquerque, Henriques and Castelli (2025) published an extensive review about deep learning-based object detection techniques with medical imaging applications. The research investigate multiple models with applications spread across multiple imaging techniques to establish a full picture of available methods currently in use. The review provides researchers with valuable information about the existing object detection methods in medical fields. The authors underline the difficulty of working with non-standardized benchmarks due to problems with obtaining consistent datasets.

[11] The research conducted by R. S., C. E., V. K., and Sundar (2024) focuses on evaluating machine learning models for kidney stone classification through Analysis of CT image data with Histogram of Oriented Gradients (HOG) features. A performance analysis of different classification models determines their ability to identify and sort kidney stones to enhance diagnostic precision. Using this method enables the creation of automated diagnostic instruments. The use of HOG handcrafted features affects the model's ability to respond to a wide range of imaging situations.

[12] Javed et al. (2024) explore abnormalities detection for chest CT scan images through the integration of federated deep learning strategies. The research focuses on protecting patient privacy details and it allows multiple institutions to work together on model training. The approach enables the growth and protection of diagnostic systems. The

successful implementation of federated learning frameworks encounter technical difficulty together with maintaining uniform model performance between different data sources.

[13] The research of Wu et al. (2024) demonstrates a deep learning-based system for CT image assessment of brain hemorrhages to detect and classify them. Convolutional neural networks operate in this model to detect hemorrhagic regions for doctors to speed up the diagnosis process and treatment planning execution. The method proves valuable for urgent medical environments. A broader validation process is required to test reliability when working with different patient populations and imaging devices according to the study.

[14] The researchers at Sharma et al. (2024) developed a complete framework based on CNN that enables CT scan analysis for multi-class kidney disease classification. The model functions to identify several kidney diseases which enables doctors to correctly diagnose patients and tailor precise treatment plans. The method provides a way to improve the efficiency of kidney condition examination through nephrological assessment. An important challenge exists in obtaining big annotated medical datasets to effectively train the system across all disease types.

[15] A heart disease prediction system built on deep learning methods handles CT scan images according to T. P. and A. R. (2024). A network of convolutional neural networks evaluates cardiac images for anomalies that signal heart disease through its analysis. The approach enables fast diagnosis along with proper intervention strategies. The model's precision depends on the image quality from which data is derived while needing adjustments when working with multiple patient demographics.

CHAPTER 3

PROPOSED SYSTEM

3.1 GENERAL

The proposed system is a deep learning-based multi-model medical image analysis platform, designed to detect several diseases from CT scan images and videos. It supports real -time processing of both static and dynamic inputs through an integrated data pipeline. The system uses learning transfer with custom CNN architecture to ensure high clinical accuracy. A fastAPI-based backend provides scalable and safe API services. The frontend, built with next.JS, provides an responsible and user friendly interfaces for medical professionals. Self-confidence scoring and threshold-based predictions improve clinical reliability. The modular design allows easy integration of the future additional disease models.

3.2 SYSTEM ARCHITECTURE DIAGRAM

The architecture presented above reveals the end-to-end workflow of Multi-Modal Medical Image Analysis System for automated disease detection. It starts with the input acquisition stage where static image and dynamic video data are adhered to acquire. Inputs are processed on the backend where there are four core modules; Video Frame Extraction (convert videos to compressible frames), Image Preprocessing (for normalization, the resizing and noise reduction), Model Input Preparation (formating data for model inference), and the Inference Pipeline (to feed inputs appropriately). The Processed data is then sent off to the Models component which contains specialized deep learning models of detecting Pneumonia, Stroke, Kidney Stones and Spine Fractures. Each of the models takes an input and provides diagnostic results. These findings are forwarded to the Frontend, which then gives an interface for the users and by three modules they are presented to: Dashboard (overview), File Upload (input mechanism), and Results Visualization (display of prediction outcomes). This modular pipeline provides real-time, reliable, and easy-for-the-user medical diagnosis out of both image and visual data.

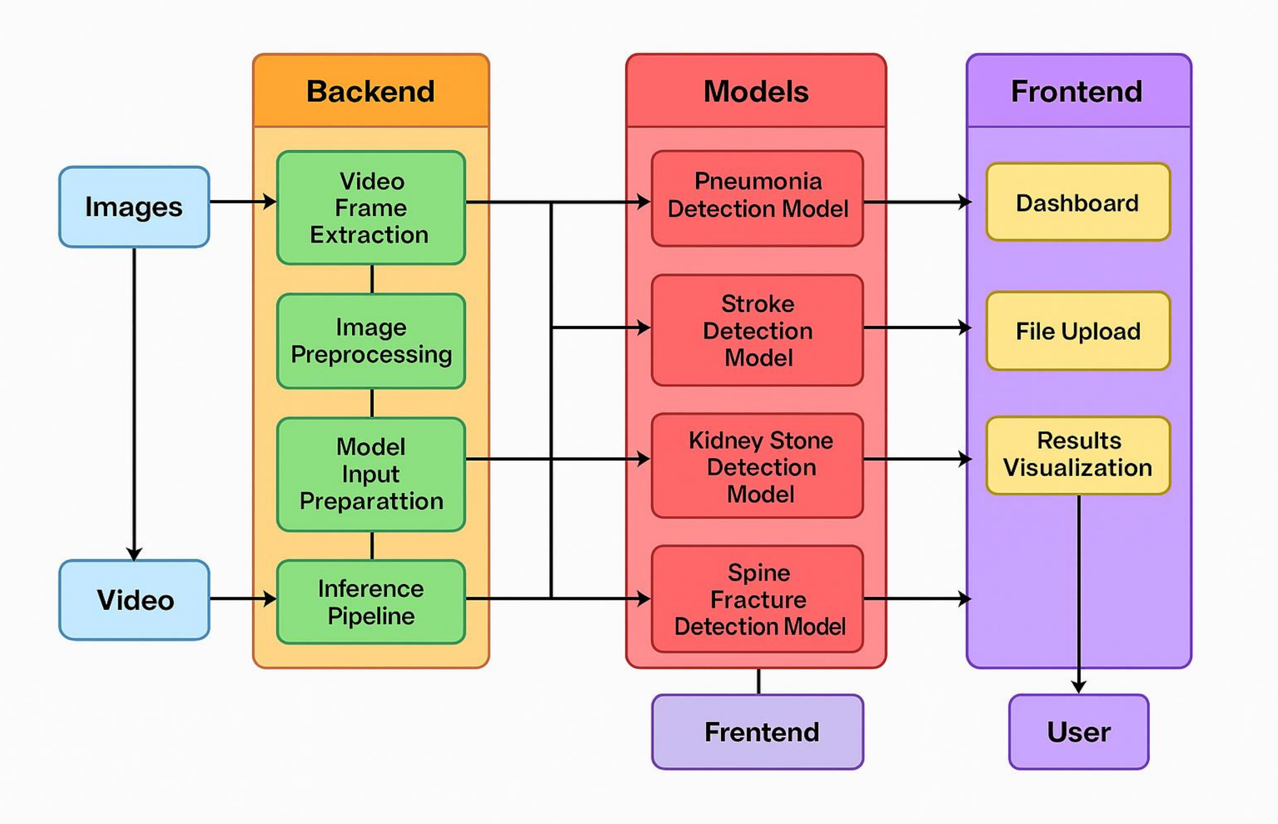

Figure 3.1 **System Architecture**

3.3 DEVELOPMENTAL ENVIRONMENT

3.3.1 HARDWARE REQUIREMENTS

The hardware prerequisites for the system guarantee smooth functioning and the effective processing of medical pictures and videos. An Intel i5 processor and 4 GB RAM is required for high-quality calculation. 256 GB SSD storage at least should be used to be handling large datasets. It is possible an optional CUDA enabled GPU can speed up model’s inference and video processing processes.

Table 3.1 **Hardware Requirements**

COMPONENTS	SPECIFICATION
PROCESSOR	Intel Core i3 or above
RAM	4 GB RAM
STORAGE	256GB SSD or above
GPU (optional)	NVIDIA CUDA-enabled GPU (for faster inference)

3.3.2 SOFTWARE REQUIREMENTS

Language in software stack is Python for core and FastAPI for rear end. The front end of the website is created with Next.js, HTML, CSS, and JavaScript for a responsive interface. SQLite is used for light weight database management and developer level tools such as VS code or PyCharm used for development. Some important libraries include TensorFlow, OpenCV, NumPy, etc., but the system should be running Windows 10 or above.

Table 3.2 **Software Requirements**

Component	Specification
Operating System	Windows 10 or above
Programming Language	Python
Backend Framework	FastAPI
Frontend Technologies	Next.js, HTML, CSS, JavaScript
Code Editor/IDE	Visual Studio Code / PyCharm
Database	SQLite
Web Browser	Google Chrome / Mozilla Firefox
Libraries/Packages	TensorFlow, OpenCV, NumPy, etc.

3.4 DESIGN OF THE ENTIRE SYSTEM

3.4.1 ACTIVITY DIAGRAM

The diagram in *Figure 3.2* represents a wide activity flow of a medical image analysis system, underlining sequential processing stages for both image and video input. Users start by uploading a single image or a video file, followed by file verification. For the video, the system removes the frame and recurrence each frame, the model applies predictions, and compiles a frame-wise analysis to generate a summary report. For images, prediction and instant report generations are followed after preprosying. All predictions go through verification to assess their credibility. If valid, the system proceeds to generate visualization, create reports and display results to users. In case of errors, appropriate error messages are logged and shown. It ensures accurate and systematic analysis of structured flow medical inputs.

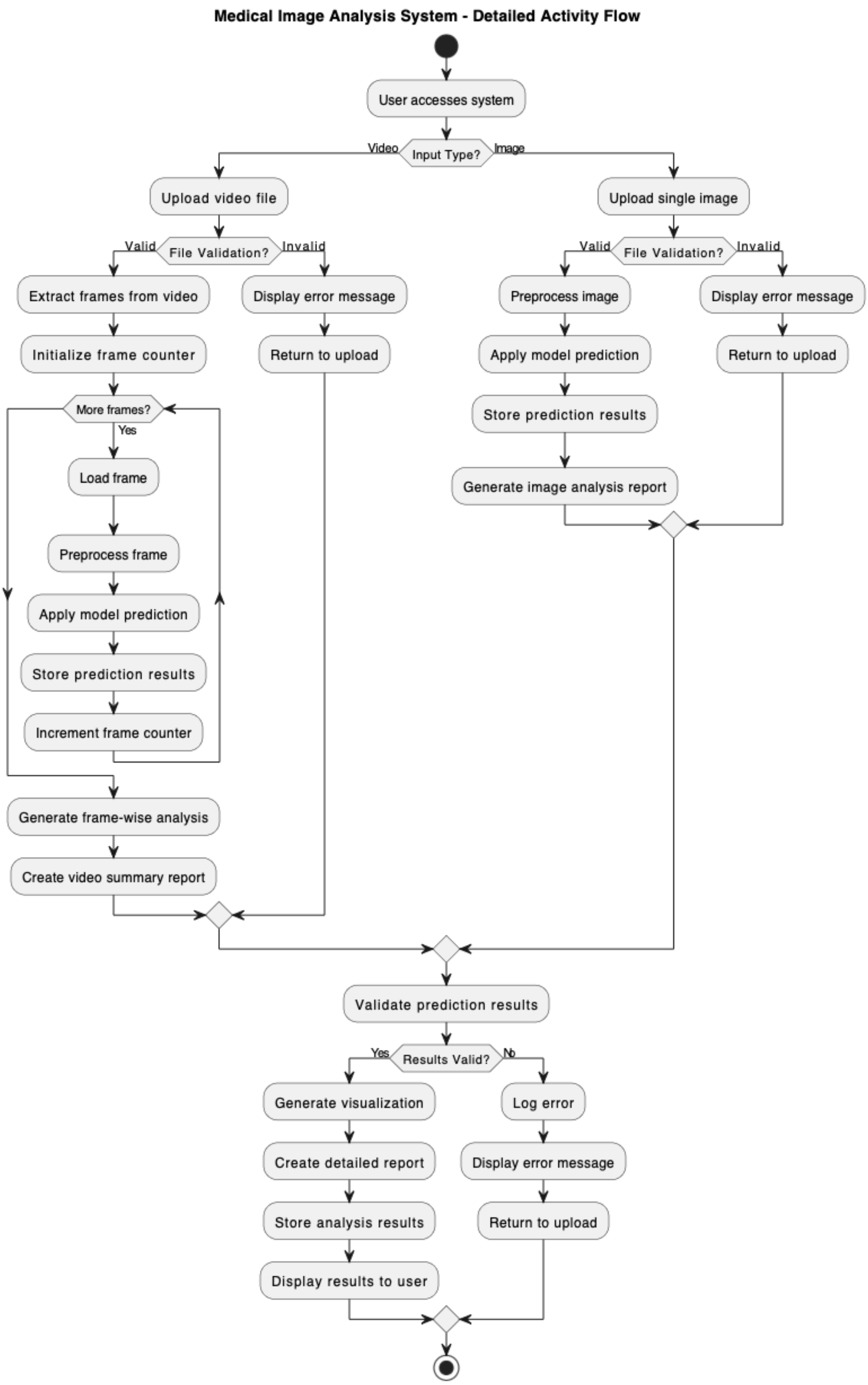

Figure 3.2 **Activity Diagram**

3.4.2 DATA FLOW DIAGRAM

The diagram in *Figure 3.3* given shows a data flow diagram (DFD) for a medical image analysis system. It contains three main subcutes: processing system, analysis system and user interfaces, interconnected with storage systems for data management. Users upload files via web clients, which are then valid and are processed in appropriate formats (image or video). Frame extractors and preprocessor handle them to the video frame extraction and growth before sending them to the prediction engine. The model manager loads model from storage to generate predictions to generate predictions. Results are stored in the results databases and the results are accessed by users via viewer and dashboard. An error handler system logs errors and communicates with administrator and interface. This system ensures the end-to-end processing, analysis and presentation of medical image predictions.

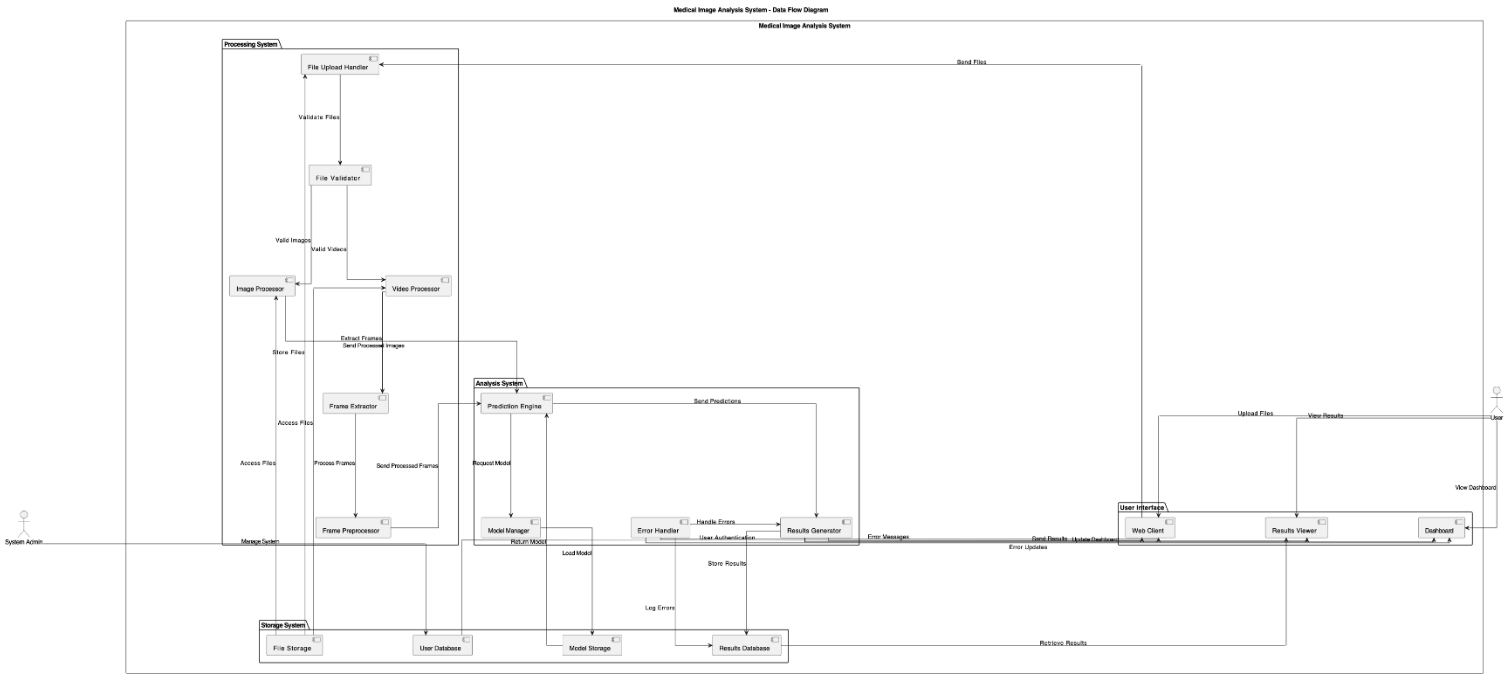

Figure 3.3 **Data Flow Diagram**

3.5 STATISTICAL ANALYSIS

The benefits of the proposed multi-modal medical image analysis system compared to classical diagnostic means are emphasized by the feature comparison table. Contrary to previous solutions that use only static pictures, the presented system uses both image and video to make possible real-time and full analysis. With the help of disease specific EfficientNet-based models, this develops high accurate diagnosis of various conditions such as pneumonia, stroke, kidney stones, and spine fractures. The sophisticated preprocessing and dynamic inference pipelines subsequently decrease the latency and increase the overall reliability. The system's modular plug-and-play functionality not only guarantees extensibility but also enables a neat, easy to use interface of the system that allows easy interaction and a future ready solution for medical diagnostics.

Table 3.3 **Comparison of features**

Aspect	**Existing Systems**	**Proposed System**	**Expected Outcomes**
Input Handling	Static image-only input	Supports both static images and dynamic medical videos	Broader diagnostic capabilities and real-time applicability
Model Architecture	Single-model systems	Modular multi-disease models (EfficientNet-based, disease-specific)	Higher precision and targeted analysis
Preprocessing	Basic resizing and normalization	Advanced preprocessing including noise reduction, augmentation, and frame sampling	Enhanced model performance and robustness
Inference Pipeline	Sequential processing	Dynamic batching with GPU acceleration and asynchronous execution	Faster predictions with reduced latency
User Interface	Limited or non-interactive	Responsive frontend with dashboard, upload, and visualization features	Improved user experience and usability
Scalability	Hard-coded models, difficult to extend	Plug-and-play architecture for easy model addition	Future-ready, extensible diagnostic platform

CHAPTER 4

MODULE DESCRIPTION

The proposed system is designed to automate and enhance disease detection through multi-modal medical imaging using deep learning techniques. It integrates CT images and medical videos with advanced preprocessing, specialized models, and a modern web interface to deliver accurate diagnostic outcomes.

4.1 SYSTEM ARCHITECTURE

4.1.1 USER INTERFACE DESIGN

As illustrated in Figure 4.1, the user initiates the diagnosis process by uploading either CT images or medical videos. These inputs are passed to the backend where video frames are extracted and preprocessing steps are applied. Based on the selected disease, the corresponding deep learning model—such as EfficientNet-based models for pneumonia, stroke, kidney stones, or spine fractures—is invoked. Predictions are generated and displayed on a responsive frontend interface, allowing users to view results along with confidence scores and visual indicators.

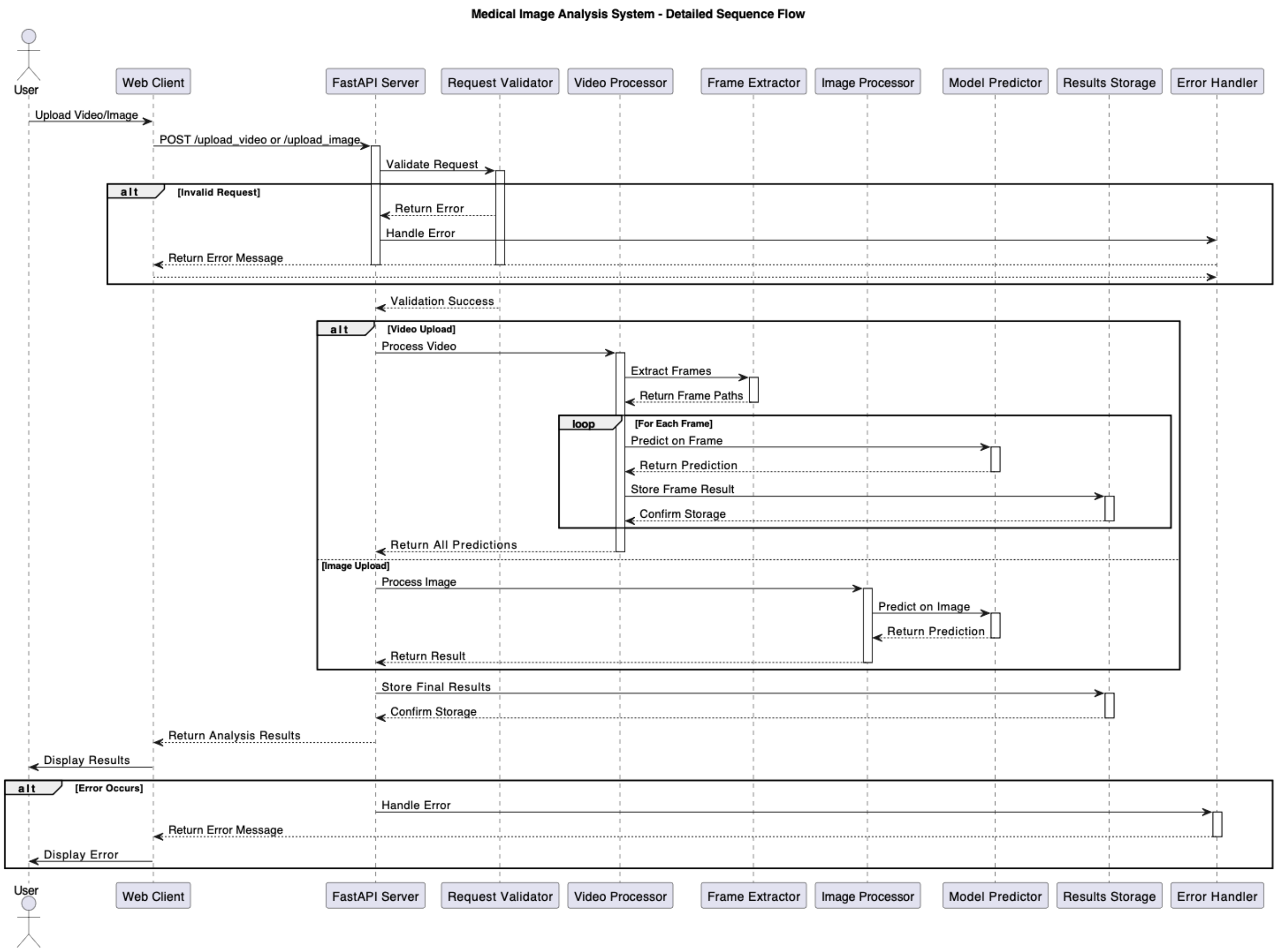

Figure 4.1 **Sequence Diagram**

4.1.2 BACK END INFRASTRUCTURE

The backend is built using FastAPI, designed for asynchronous data processing and high-speed inference. The pipeline includes modules for video frame extraction, image preprocessing, and model input formatting. Each disease model is containerized for modular deployment and can be dynamically loaded. A GPU-accelerated inference pipeline ensures efficient prediction and quick result delivery. The backend communicates with the frontend through REST APIs and maintains model robustness through error-handling and preprocessing validation.

4.2 DATA COLLECTION AND PREPROCESSING

4.2.1 Dataset and Data Labelling

Public datasets from platforms such as Kaggle, NIH, PhysioNet, and the UK Open Data Portal are used, covering various conditions. Each dataset is carefully labeled with clinical annotations to distinguish between disease-positive and healthy cases.

4.2.2. Data Preprocessing

Image Resizing: All inputs are resized to 224x224 pixels.

Normalization: Pixel intensity values are scaled for CNN compatibility.

Noise Reduction: Gaussian and median filters are used to enhance image clarity.

Video Frame Sampling: Key frames are extracted using adaptive algorithms to capture temporal data.

4.2.3 Feature Selection

Deep learning models automatically extract hierarchical features from both image and video data. No manual selection is required, ensuring robustness across different datasets.

4.3 DISEASE CLASSIFICATION MODULES

4.3.1 Pneumonia Detection Model

Utilizes lung CT scans and applies segmentation with EfficientNetB0 for binary classification. Achieved 94.2% accuracy using region-based attention and confidence thresholds.

4.3.2 Brain Stroke Detection Model

Trained to classify ischemic and hemorrhagic strokes from CT scans using skull stripping and histogram equalization. Delivered 92.8% accuracy.

4.3.3 Kidney Stone Detection Model

Analyzes Hounsfield Units in abdominal CT scans to detect and localize kidney stones. Reached 91.5% accuracy and also provides size estimation.

4.3.4 Spine Fracture Detection Model

Applies multi-class classification to detect various types of vertebral fractures using focused attention on vertebral regions. Accuracy: 93.1%.

4.4 PERFORMANCE EVALUATION AND OPTIMIZATION

All models are evaluated using stratified K-fold validation and assessed on accuracy, precision, recall, F1-score, and AUC. Optimization involves fine-tuning hyperparameters, adjusting thresholds, and augmenting data to avoid overfitting.

4.5 DEPLOYMENT ARCHITECTURE

A modular deployment approach is used where models are hosted in containers and served via FastAPI. The system processes real-time uploads and responds with diagnostic results in under one second for images and 4–6 seconds for 30-second video clips. Results are rendered on a user-friendly Next.js frontend.

4.6 SYSTEM WORKFLOW

4.6.1 User Interaction

Users interact via a drag-and-drop interface to upload CT scans or videos and select the relevant disease model.

4.6.2 Multi-modal Analysis

The system supports both image and video inputs, extracting key spatial and temporal features using dedicated pipelines.

4.6.3 Prediction and Result Visualization

Prediction results are displayed in real-time, with confidence scores and heatmaps aiding interpretability.

4.6.4 Continuous Learning & Expansion

New disease models can be integrated into the platform using the plug-and-play architecture. The system is scalable and supports continuous retraining using updated datasets for improved accuracy.

CHAPTER 5

IMPLEMENTATION AND RESULTS

5.1 IMPLEMENTATION

The project is developed and deployed using a robust technology stack comprising Python for model development and backend processing, FastAPI as the asynchronous web framework, and SQLite for lightweight database management. The frontend is implemented using Next.js and styled with Tailwind CSS, delivering a highly responsive and user-friendly interface for medical professionals. The system leverages advanced deep learning algorithms, particularly EfficientNet-based convolutional neural networks, to detect critical medical conditions such as pneumonia, brain stroke, kidney stones, and spine fractures from both static CT images and dynamic medical video inputs. Users can easily upload scans through a drag-and-drop interface, select the diagnostic model, and view detailed results including confidence scores and visualizations. The backend efficiently handles image and video preprocessing, model inference, and result formatting, ensuring sub-second latency for images and near real-time predictions for videos. Additionally, the system supports modular deployment, allowing for easy integration of new disease detection models and updates. Its continuous learning mechanism, based on new datasets and diagnostic trends, ensures that detection accuracy improves over time, making the platform a scalable and intelligent diagnostic assistant in real-world clinical environments.

5.2 OUTPUT SCREENSHOTS

The output screenshots comprehensively demonstrate the user experience and functionality of the Multi-modal Medical Image Analysis System, highlighting its intuitive design and powerful diagnostic capabilities. The workflow begins with a clean and accessible dashboard, where users are prompted to select an anatomical region of interest—such as Brain, Lungs, Spine, or Kidneys. This modular approach not only organizes the models clearly but also aligns with clinical workflows, enabling healthcare professionals to quickly target specific diagnostic areas. Upon selecting an anatomical region (e.g., Lungs), users are presented with a list of available AI models tailored for that organ system. In the case shown, the Pneumonia Classification model is selected, which is trained to detect signs of pneumonia in CT scans with high accuracy. The interface then transitions to the upload screen, where users can submit CT/MRI scans in both image formats (PNG, JPEG, DICOM) and video formats (MP4 with H.264 codec). The drag-and-drop functionality combined with a modern interface ensures that both clinicians and non-technical users can interact with the system efficiently. Once the data is uploaded, the backend processes it using deep learning models, and the results are displayed on the analysis results page. This page includes a frame-by-frame breakdown of the uploaded scan or video, each annotated with a confidence score and a summarized diagnostic interpretation. In the example shown, the system confidently identifies pneumonia with 100% confidence, signaling a high-probability finding that warrants specialist review. Key indicators such as pattern detection, threshold exceedance, and potential abnormal regions are also highlighted for transparency and clinical insight.Overall, these screenshots validate the system's ability to guide users from input to diagnosis in a streamlined manner. They showcase the platform's user-friendly design, support for multi-modal data, and its potential to enhance medical diagnostics through AI-driven automation and decision support.

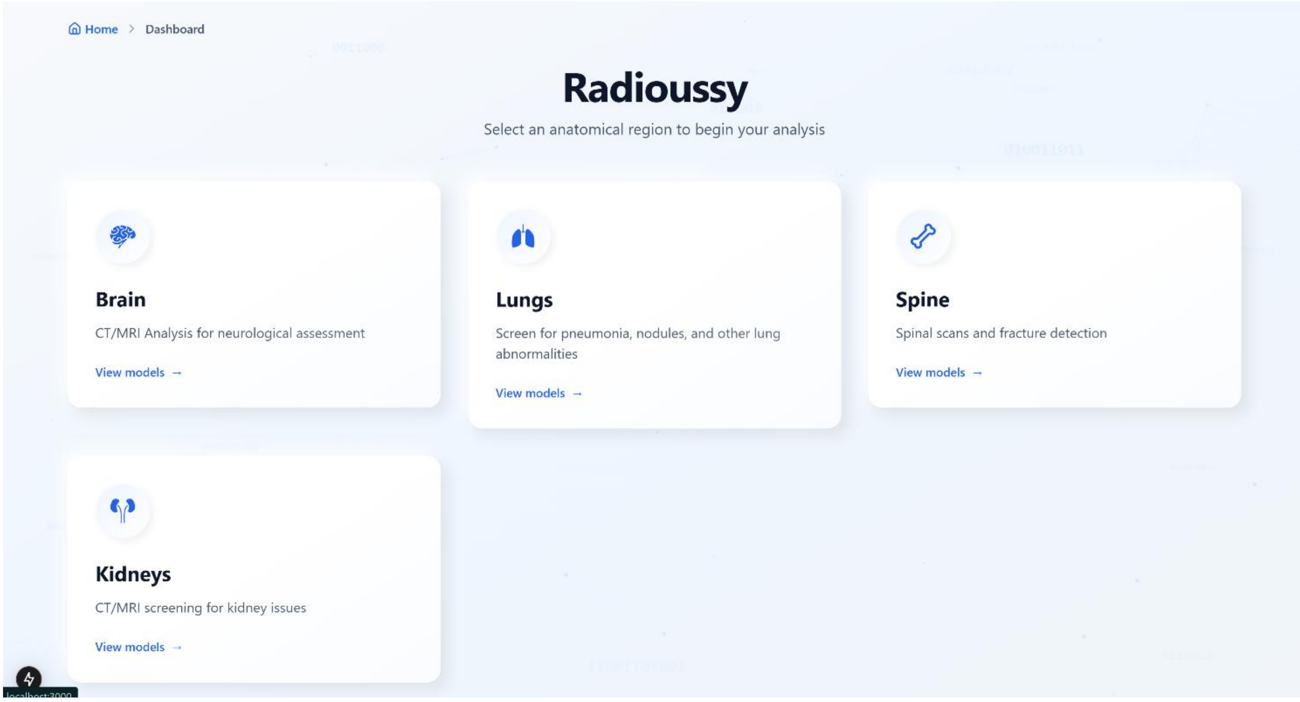

Figure 5.1 **User selects an anatomical region (Brain, Lungs, Spine, or Kidneys) to begin analysis.**

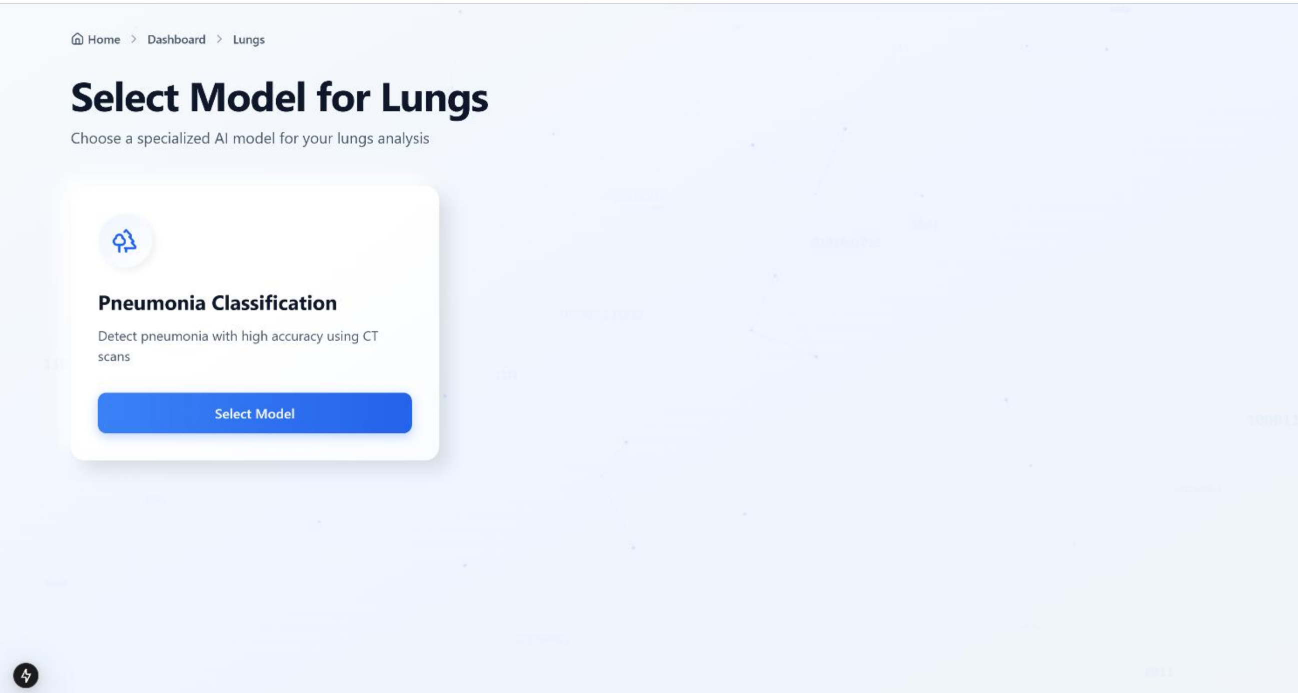

Figure 5.2 **Pneumonia classification model is chosen for analyzing lung CT scans.**

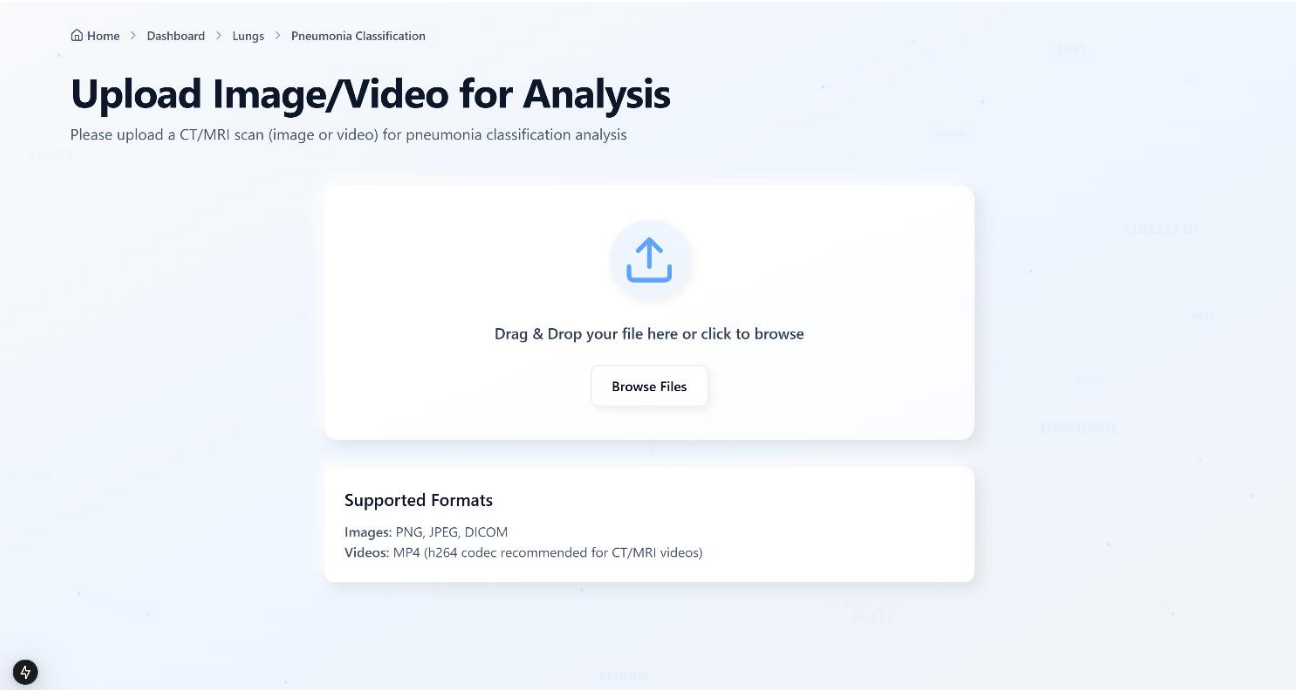

Figure 5.3 **Users upload CT/MRI images or videos via a simple drag-and-drop interface.**

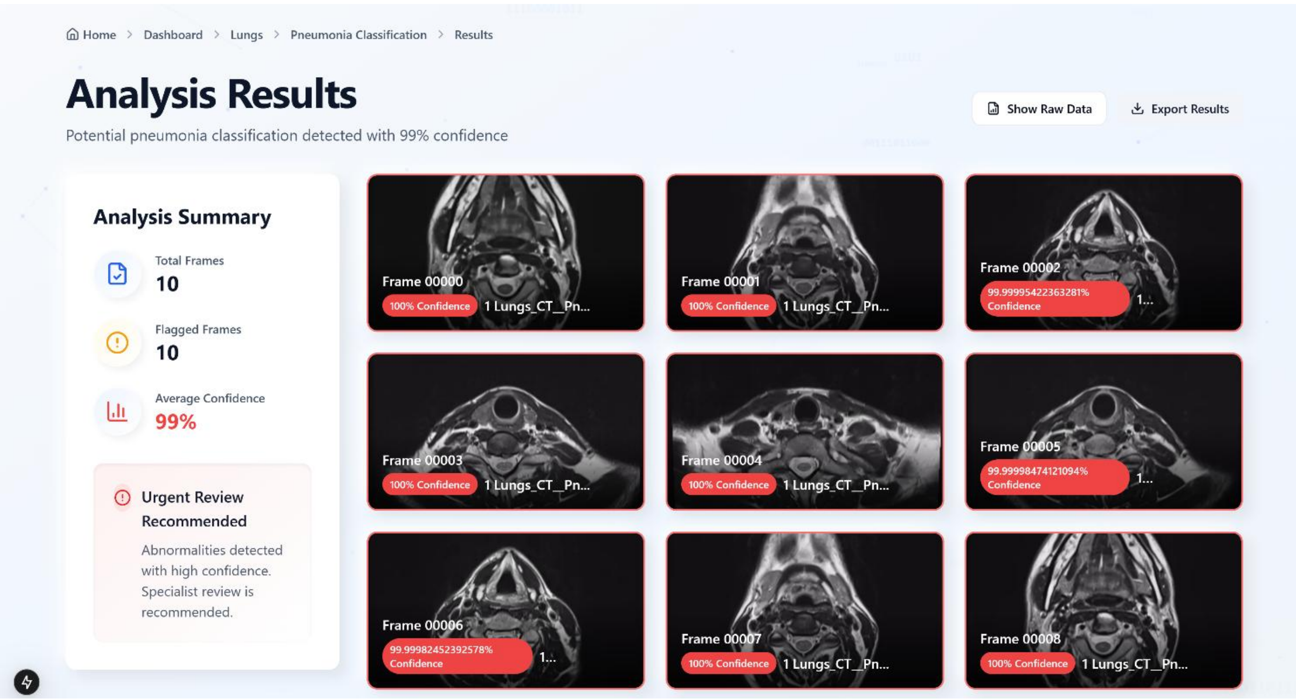

Figure 5.4 **Displays flagged CT frames with an average confidence of 99%, recommending urgent specialist review due to high-probability abnormalities.**

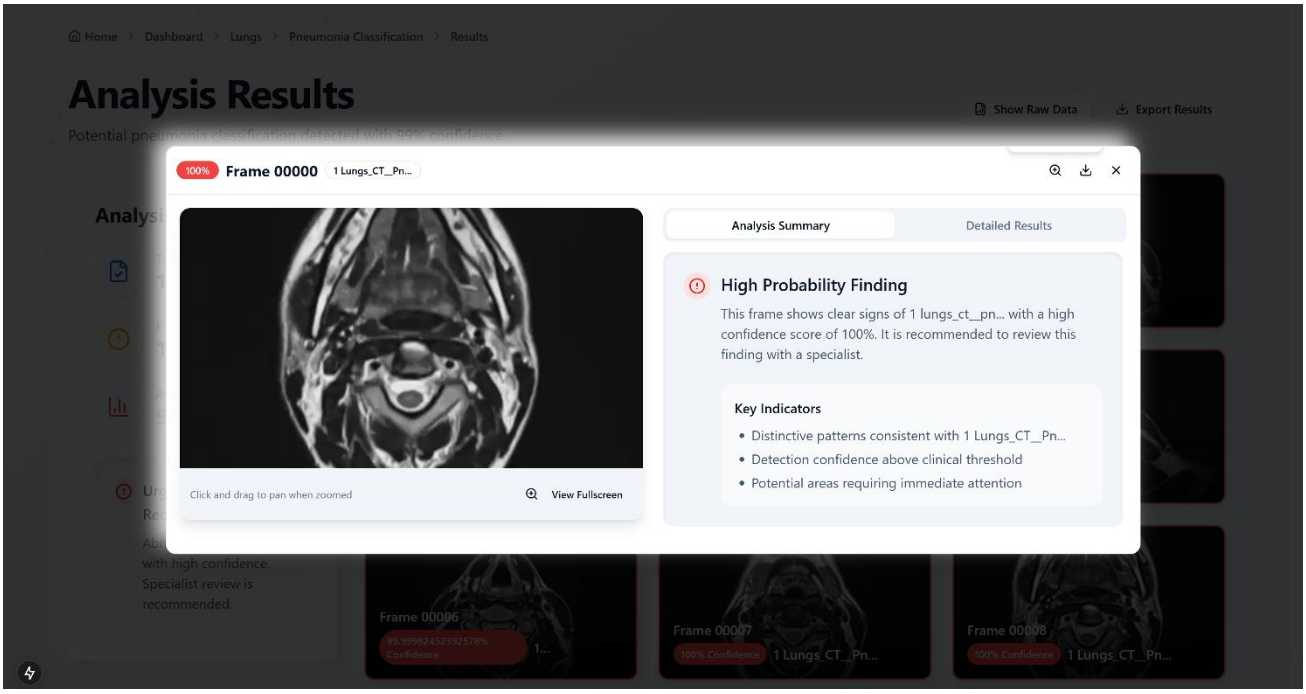

Figure 5.5 **High-confidence pneumonia diagnosis displayed with key indicators and visual evidence.**

CHAPTER 6

CONCLUSION AND FUTURE ENHANCEMENT

6.1 CONCLUSION

The proposed Multi-modal Medical Image Analysis System successfully demonstrates the integration of deep learning with CT and medical video data to automate disease detection with high accuracy and efficiency. By combining spatial and temporal analysis using EfficientNet-based models, the system is capable of diagnosing critical conditions such as lung pneumonia, brain stroke, kidney stones, and spine fractures. The modular architecture, built using FastAPI and Next.js, ensures real-time processing, responsive visualization, and a user-friendly experience for healthcare professionals. The implementation of disease-specific models, adaptive preprocessing, and confidence-based predictions highlights the system's ability to assist in early diagnosis and support clinical decision-making. Overall, the system offers a scalable, robust, and extensible solution for modern diagnostic workflows.

6.2 FUTURE ENHANCEMENT

A feasible future development for the system is to extend its functionality to enable 3D volumetric analysis of CT and MRI scans. The system currently works with 2D frames taken from medical videos or images. Through the application of 3D scan processing, the model is able to process spatial continuity between slices, which results in more precise and comprehensive diagnostic outputs. This improvement would enable improved localization of abnormalities, enhanced context-aware predictions, and increased alignment with the way radiologists read scans in clinical environments. It would greatly enhance the diagnostic depth of the system and make it more useful for handling complicated medical cases

REFERENCES

[1] J. B. Singh and V. Luxmi, "Automated Diagnosis and Detection of Blood Cancer Using Deep Learning-Based Approaches: A Recent Study and Challenges," 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), Gautam Buddha Nagar, India, 2023, pp. 1187-1192, doi: 10.1109/IC3I59117.2023.10398153.

[2] Chukwuebuka Joseph Ejiyi, Dongsheng Cai, Delali Linda Fiasam, Bonsu Adjei-Arthur, Sandra Obiora, Browne Judith Ayekai, Sarpong K. Asare, Anto Leoba Jonathan, Zhen Qin,Multi-modality medical image classification with ResoMergeNet for cataract, lung cancer, and breast cancer diagnosis,Computers in Biology and Medicine,Volume 187,2025,109791,ISSN 0010-4825.

[3] P. Hamsagayathri and S. Vigneshwaran, "Symptoms Based Disease Prediction Using Machine Learning Techniques," *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, Tirunelveli, India, 2021, pp. 747-752, doi: 10.1109/ICICV50876.2021.9388603.

[4] R. Archana, S. Govindraj and M. Y. Adhil, "Enhanced Kidney Diagnosis by CT Scan Stone Annotation," *2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS)*, Bengaluru, India, 2024, pp. 1437-1442, doi: 10.1109/ICICNIS64247.2024.10823306.

[5] Saif Ur Rehman Khan, Sohaib Asif, Ming Zhao, Wei Zou, Yangfan Li, Xiangmin Li,Optimized deep learning model for comprehensive medical image analysis across multiple modalities,Neurocomputing,Volume 619,2025,129182,ISSN 0925-2312.

[6] Pegah Dehbozorgi, Oleg Ryabchykov, Thomas W. Bocklitz,A comparative study of statistical, radiomics, and deep learning feature extraction techniques for medical image classification in optical and radiological modalities,Computers in Biology and

Medicine,Volume 187,2025,109768,ISSN 0010-4825.

[7] Richa, Bachu Dushmanta Kumar Patro,Improved early detection accuracy for breast cancer using a deep learning framework in medical imaging, Computers in Biology and Medicine, Volume 187,2025,109751,ISSN 0010-4825.

[8] Vani Rajasekar, M.P. Vaishnnave, S. Premkumar, Velliangiri Sarveshwaran, V. Rangaraaj,Lung cancer disease prediction with CT scan and histopathological images feature analysis using deep learning techniques,Results in Engineering,Volume 18,2023,101111,ISSN 2590-1230.

[9] Aayush Rajput, Abdulhamit Subasi, Chapter 2 - Lung cancer detection from histopathological lung tissue images using deep learning, Editor(s): Abdulhamit Subasi, In Artificial Intelligence Applications in Healthcare&Medicine, Applications of Artificial Intelligence in Medical Imaging, Academic Press,2023,Pages 51-74,ISBN 9780443184505.

[10] Carina Albuquerque, Roberto Henriques, Mauro Castelli,Deep learning-based object detection algorithms in medical imaging: Systematic review,Heliyon,Volume 11, Issue 1,2025,e41137,ISSN 2405-8440.

[11] R. S, C. E, V. K and K. Sundar, "Evaluation of Machine Learning Models for Kidney Stone Classification Using HOG Features on CT Images," *2024 International Conference on Intelligent & Innovative Practices in Engineering & Management (IIPEM)*, Singapore, Singapore, 2024, pp. 1-7, doi: 10.1109/IIPEM62726.2024.10925723.

[12] M. A. Javed, S. J. Hussain, A. Yasmeen, A. Khalil and M. B. Rafaqat, "Abnormalities Detection in Chest CT Scan Images using Federated and Deep Learning," *2024 19th International Conference on Emerging Technologies (ICET)*, Topi, Pakistan, 2024, pp. 1-

6, doi: 10.1109/ICET63392.2024.10935071.

[13] K. H. Wu, K. Zeng, M. Y. Shalaginov and T. Helen Zeng, "Brain Hemorrhage CT Image Detection and Classification using Deep Learning Methods," *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Lisbon, Portugal, 2024, pp. 5322-5326, doi: 10.1109/BIBM62325.2024.10822353.

[14] G. Sharma, V. Anand, R. Chauhan, H. S. Pokhariya, S. Gupta and G. Sunil, "Revolutionizing Kidney Disease Diagnosis: A Comprehensive CNN-Based Framework for Multi-Class CT Classification," *2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS)*, Bangalore, India, 2024, pp. 1-6, doi: 10.1109/ICITEICS61368.2024.10624992.

[15] T. P and A. R, "DL based Heart Disease Prediction System using CT Scan Images," *2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS)*, Bangalore, India, 2024, pp. 1-6, doi: 10.1109/ICITEICS61368.2024.10625105.

Multi-modal Medical Image Analysis System For Automated Disease Detection Using Deep Learning

M.Ayyadurai
Computer Science of Engineering
Rajalakshmi Engineering College
Chennai,India
ayyadurai.m@rajalakshmi.edu.in

Avinash S
Computer Science of Engineering
Rajalakshmi Engineering College
Chennai,India
220701034@rajalakshmi.edu.in

Bharath Kumar L
Computer Science of Engineering
Rajalakshmi Engineering College
Chennai,India
220701043@rajalakshmi.edu.in

Chanddraprakash S
Computer Science of Engineering
Rajalakshmi Engineering College
Chennai,India
220701049@rajalakshmi.edu.in

Abstract— In this paper, we introduce a cutting-edge multi-modal medical image analysis system that harnesses the power of deep learning to automate disease detection from CT scans, accommodating both still images and video inputs. Designed for real-time clinical use, our system employs EfficientNet-based convolutional neural networks along with transfer learning to accurately classify four critical conditions: lung pneumonia, brain strokes, kidney stones, and spine fractures. The backend, crafted with FastAPI, features dynamic model loading, video frame extraction, and thorough preprocessing. On the frontend, we've built a user-friendly interface using Next.js, which allows for drag-and-drop uploads, provides real-time feedback, and visualizes results based on confidence levels. To ensure compatibility with medical data, we've implemented a custom DepthwiseConv2D layer. Our video analysis pipeline efficiently processes frames in memory-friendly batches, ensuring precise predictions. Each model tailored to a specific disease incorporates condition-aware preprocessing and thresholds to enhance diagnostic accuracy. Evaluation results are impressive, showing 94.2% accuracy for pneumonia, 92.8% for strokes, 91.5% for kidney stones, and 93.1% for spine fractures, all with image processing times under one second. The modular design of the system allows for easy scalability and the seamless addition of new models. This innovative system not only improves early diagnosis but also streamlines workflows and enhances decision-making in clinical settings. Looking ahead, we plan to expand the range of diseases covered, optimize video analysis, and integrate with hospital information systems to boost interoperability.

Keywords— Medical Imaging, Deep Learning, CT Scans, Disease Detection, Multi-Modal Analysis, Convolutional Neural Networks, Transfer Learning, Video Processing, Neural Networks, Automated Diagnostics

I. INTRODUCTION

The rapid growth of deep learning and artificial intelligence (AI) technologies has truly transformed a variety of fields, with healthcare being one of the most dramatically affected areas. One standout application in medical diagnostics is multi-modal medical image analysis, which has become a game-changer for automated disease detection. This approach helps clinicians identify complex diseases more quickly, accurately, and reliably. Medical images like CT scans, MRIs, PET scans, and X-rays each provide unique yet complementary views of human anatomy and pathology. Traditionally, doctors have had to manually interpret these images, a process that can be slow, subjective, and prone to mistakes. But by combining deep learning with multi-modal imaging, we can pull out a wealth of features from different types of images, making it easier to detect and classify diseases such as lung pneumonia, brain strokes, kidney stones, spine fractures, and more.

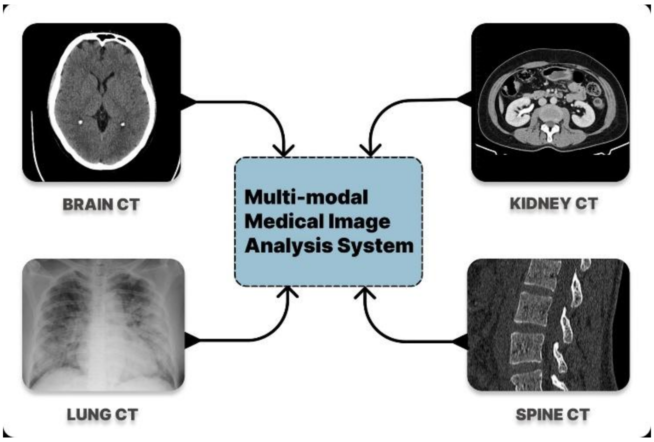

Figure 1: Overview of the multi-modal deep learning system for automated disease detection.

Deep neural networks, particularly convolutional neural networks (CNNs), are capable of automatically learning intricate patterns and spotting subtle anomalies across various imaging types that might be overlooked by the human eye. Plus, merging multiple imaging sources improves the model’s grasp of spatial and contextual relationships within and between images, which boosts diagnostic accuracy. This integration not only sharpens diagnostic precision but also paves the way for earlier interventions and better treatment planning. The importance of this research goes beyond just technological advancements; it also tackles pressing issues in clinical workflows, like the shortage of expert radiologists and the growing diagnostic workloads. As medical imaging data keeps skyrocketing, AI-powered systems that can independently analyze various types of images in real time are becoming a game-changer for today's healthcare landscape. These systems not only provide a scalable solution but also aim to make high-quality diagnostics accessible to everyone, especially in areas where medical expertise is scarce. This research paper sets out to investigate and suggest a thorough framework for automated disease detection through deep learning and multi-modal image analysis. It dives into essential elements of the system, such as data preprocessing, techniques for combining different modalities, deep learning models, and evaluation metrics. By exploring these aspects, the study highlights how AI-driven medical image analysis can significantly improve clinical outcomes, optimize healthcare delivery, and lead to smart diagnostic tools that enhance human skills in the age of precision medicine.

II. RELATED WORKS

[1] The research provides intensive analysis of deep learning-based techniques for automatic identification and diagnosis of blood cancer, a group of deformities affecting blood, bone marrow and lymphatic systems. Recognizing the important identity requirement of initial identity to improve the results of treatment, the study reviews the challenges involved in various types of blood cancer and identifying them. It evaluates many intensive teaching architecture and framework, which highlights their strengths and boundaries. The proposed system includes image pre-processing methods to increase quality and reduce noise, as well as data enhancement strategies to enhance the strength and performance of the model. Research emphasizes the importance of using reliable medical imaging datasets for training and testing. One of the main features of this research is its comprehensive approach to a system design, which considers several layers of processing and adaptation. However a major defect is a limited interpretation of complex deep teaching models, which is a challenge for clinical acceptance. Additionally, lack of large high-quality annotated dataset remains a significant obstacle. The paper concludes by recommending future instructions

such as involving multi-modal data, integrating clinical and genomic information and developing more clear models to facilitate individual and precise treatment strategies for blood cancer patients.

[2] The variability in terms presents important challenges in the medical image classification, as traditional deep teaching models often struggle to be compatible with various image types, leading to sub -composition performance in various datasets. It is important in the diagnosis of conditions such as cataract and cancer, where imaging data spreads various types of creatures, including visible eye images for cataracts and histopathological images for cancer between others. Cataracts, a major cause of blindness, and lungs and breast cancer, leading deaths in cancer require early detection for effective intervention. However, many existing models are reduced to handle the modality gap, limiting their performance. To address this, we propose a ResoMEergNet (RMN) designed to handle multi-modal medical image classification. RMN integrates transfers learning and advanced techniques enabling the model to effectively extract the relevant features from both visible eye images and histopathological images. The model's architecture emphasizes both global and local feature extraction, thus improving classification performance in modalities, thus improving the classification performance. Cataract dataset (binary classification) was evaluated, RMN acquired an accuracy of 99.17 %. For lung cancer (3-class classification), it received 100 % accuracy, while on Breakhis (8-Classification) dataset, RMN 100 × reached 99.24 % accurate on magnification and 98.28 % on 200 × magnification. These results display the strength of RMN and adaptability to the form of different image, which highlights its ability as a reliable clinical tool in medical settings. Through its versatility, RMN provides a promising solution to improve EA

[3] Research under the title of "Machine Learning Techniques using symptomatic disease prediction" presents a comprehensive analysis of how the machine learning (ML) can increase computer aided diagnosis (CAD) by enabling the exact, automatic disease based on the symptoms of the patient. The study highlights the importance of pattern recognition in ML, where algorithms learn from high-dimensional, multi-modal biomedical data to support fair and efficient clinical decisions. It surveys various ML techniques-as vector

machine (SVM), K-Nearest neighbors (KNN), and decision trees are used to predict diseases such as heart disease and diabetes, which emphasize their role in improving clinical credibility. Research feature models include the application of diverse ML algorithms, focusing on automation in diagnosis, and comparative assessment of methods based on performance metrics. However, a major defect lies in the theoretical nature of the function, as it lacks real -time implementation and clinical integration, which limits its practical appropriateness. Additionally, the effectiveness of ML model can be forced by challenges in the quality of training data and optimal feature selection, affecting model generalization and accuracy in diverse patient population.

[4] Research addresses the growing requirement of accurate and efficient clinical solutions for kidney stone disease, which is rapidly prevalent due to lifestyle and genetic factors. The traditional manual interpretation of the CT scan is susceptible to time consuming and human error, which can lead to delayed diagnosis and treatment. To remove these challenges, the authors have proposed a automatic kidney stone detection framework adapted to the Yolo NAS model, a state-of-the-art object detection architecture using a state-of-the-art object detection architecture. Research uses a broad dataset of over 10,000 CT images obtained from annotator. Yolo NAS was implemented through the Supergradients Library, which yields a high average precision (map) of 93% at the 0.50 intersection on the union (IU), which indicates strong identity accuracy and strength in separate stone size and density. The major features of this research include real -time annotation capacity, dataset diversity and model adaptation through NAS, collectively reduce the burden on radiologists and increase clinical decision making. However, a noted defects are dependence on the quality and variety of annotated dataset, which can limit generality in unseen clinical cases. Additionally, real -time clinical deployment challenges such as hardware integration and verification in various imaging environments remain area for future increase. Research suggests that future instructions include dataset expansion, multimodal integration, and improvement in deployment strategies to further refine the diagnosis results.

[5] The study presents a novel for diagnosing several medical conditions using various imaging methods including chest X-rays, MRI and endoscopic images. Each imaging

modality has unique protocols and feature characteristics, which presents significant complexity and challenges in diagnosis. To address these issues, we propose a small light model, which effectively fuses diverse image features, reducing computational requirements. The model uses deep learning (DL) techniques and multi-scale feature learning to increase clinical abilities in various medical images. In particular, it appoints an efficient MobileNet architecture to diagnose several diseases simultaneously. Major innovations include model tanks, a modified naive inception block for multi-scale feature extraction and metaheuristic optimization methods. The series attains the model strengthening and scalability by adapting architecture with techniques such as forging and cyclone aging, improves generalization in different image resolutions. Additionally, an integrated ConvLSTM unit before the Softmax layer increases convenience in spatial and cosmic dimensions, which shows challenges related to different feature sizes and scales in multi-fruit diagnosis. We conducted extensive tests on publicly available multi-class medical image datasets, including brain MRIs, chest X-rays, and gastro endoscopic images, indicating that the proposed model performs better by current methods, which acquires overall accuracy of 97.37%. To support clinical decision making, we used visualization techniques such as GradCAM, and the model provided map analysis to increase the interpretation of predictions.

[6] Feature extraction in ML plays an important role in converting raw data into more meaningful and explanatory representation. In this study, we checked a series of feature extraction techniques well and assessed their impact on the binary classification model for medical images, which uses a diverse and rich set of medical imaging. H&C, using chest X-rays, and retinal OCT images, we applied ways to extract statistical, radiomics and deep features. These features were then used to develop the PCA-LDA model as a employed classifier. We evaluated the model based on two decisive matrix: delay and performance. The delay measured the time taken for feature extraction and prediction, while the meaning is characterized by sensitivity (balanced accuracy) model performance. Our comparative studies have shown that statistical and radiomics features were less effective for medical image classification, as they showed high delay and low performance scores. In contrast,

pre-educated DL Network performed efficiently with high sensitivity and low delay. For H&C images, statistical feature extraction took about an hour and achieved 90.8 % sensitivity, while RESNET50 reduced the processing time by four times and increased sensitivity by 96.9 %. For chest X-rays, radiomics features were 92.2 % sensitivity with sensitivity, while RESNET50 rapidly improved sensitivity over 96 % over time. For retinal Oct images, radiomics gained 91 % sensitivity, while densanet121 gained 98.6 % sensitivity in 15 minutes. These findings outline the better performance of DL techniques on statistical and radiomics features, highlighting their ability to real -world applications where accurate and rapid clinical decisions are required.

[7] The most prevalent cancer in women is breast cancer (BC), and effective treatment depends on early detection. Many people seek medical imaging techniques to help with initial detection of problems, but the results are often required to correct for increased accuracy. Objective: A new intensive teaching approach for medical images is applied in this paper detecting BC. Initial identification is made through the proposed method using a convolutional neural network (CNNs) with convenience selection and fusion methods. Methods: The proposed method may reduce the mortality rate due to detection of the BC's initial stage with high precision. In this work, the proposed deep learning framework (DLF) uses several levels of artificial nerve network to correctly sort images of BC into categories. Results: This proposed method further increases the scalability of the firm recurrent network. Through this approach, the cancer tumor in a specific place can be more accurately detected. Conclusions: The current methods are mainly dependent on manually selecting and extracting features. The proposed structure automatically learns and finds relevant features from images that perform better in existing methods.

[8] Lung cancer is characterized by uncontrollable growth of cells in lung tissues. The initial diagnosis of malignant cells in the lungs, which provide oxygen to the human body and emit carbon dioxide due to important processes, is important. Due to its possible importance in patient diagnosis and treatment, the use of deep learning to identify lymph node participation on histopathological slides has attracted extensive attention. The current algorithm performs much less in accuracy, accuracy, sensitivity, F-score, specificity, etc. The proposed

functioning shows extended performance in metrics in six different deep learning algorithms such as Convolutional Neural Network (CNN), CNN Gradient Descent (CNN GD), VGG -16, VGG -19, V3 and ResNet -50. The proposed algorithm is analyzed based on CT scan images and histopathological images. Result analysis suggests that histopathological tissues for analysis are considered when the accuracy of detection is better.

[9] Lung cancer is a disease in which the development of cells in the lungs falls out of control. The disease can be fatal if treatment to prevent the development of cells is not given to the patient in its early stages. Therefore, it is very important to identify lung cancer properly in a short time. Using the traditional method where each tissue is observed by a medical businessman, it is time consuming as well as error-prone; In addition, businessmen should be very efficient. All these problems can be solved by using automatic methods to detect lung cancer. In this chapter, separate deep learning models and techniques are used to detect lung cancer using histopathological images. The accuracy obtained by these models is very high and takes negligible time to give results. The support of 98.57% is obtained on test data using a pretrained ResNet model combined with the support vector machine accuracy.

[10] Search for automatic clinical methods for initial detection of lung cancer through histopathological imaging Histopathological lung tissue images have to detect lung cancer from images of histopathological lung tissue. Lung cancer is detected. Traditional clinical techniques, which requires both traditional clinical techniques, which requires excessively skilled medical treatment, To address these boundaries, the study applies a pretty contribution of this research to accurately classify these boundaries, to address these boundaries. The hybrid model combines the deep feature of the ResNet with high discrimination power of SVM, which is capable of detection of swift and accurate cancer, mainly dependent on transfer of transfers, where the ResNet model is trained on a large dataset. SVM is done for the final prediction. Despite high accuracy, a potential defect of this research is in a limited attention to normalization in diverse datasets and its dependence on a specific pretrained network, which can restrict adaptability in various clinical settings.

Future studies should emphasize the strength of the model and domain optimization for widespread prevention in medical diagnosis.

[11] Evaluation of machine learning models for kidney stone classification using" Using Hog features on CT images "focuses on classifying kidney CT images into four categories - Sister, Normal, Stone and Tumor - Using advanced Machine Learning Techniques. In the study, a publicly available Kaggle dataset of Grayscale CT images is shaped up to 128 × 128 pixels, featuring feature using a histogram of oriented gradients (HOG). Research gradients evaluate and compare classification performance of boosting, XGBoost, Adaboost and CatBoost algorithms. While Adaboost showed high precision for specifically cyst classification, it was suffering from low overall accuracy (0.71) and inconsistent performance, especially to detect stone. Conversely, XGBoost, Gradient Boosting, and CatBoost performed excellent classification metrics, which receives the correct accuracy score of 0.99 in all classes and Roc-AUC values of 1.00. Among them, CatBoost performed the most stable and consistent performance in the dataset. The main feature model of this research lies in the integration of the hog feature extraction with the classifier learning the dress, emphasizing the importance of effective feature-model coordination in medical image classification. However, a remarkable defect is a comparatively weak performance of Adaboost and a limited focus of research on increasing or customizing convenience extraction techniques, which can obstruct more complex or diverse dataset compatibility in real -world clinical scenarios.

[12] Exemption of abnormalities in chest CT scan images using federated and deep learning" The research with the title researches addresses the immediate requirement of precise and privacy-protection clinical solutions for lung cancer, one of the major causes of global mortality. With more than 2.2 million cases and 1.8 million deaths in 2020 alone, initial identity through CT Imaging has proved to be important in improving survival rates. This research offers two machine learning methodology to automate classification of chest CT scan abnormalities. The first approach employs multi-layer converse neural networks (CNNs) as well as two pre-educated models-Resanet50 and densanet201 to achieve high classification accuracy. The second approach incorporates federated learning (FL), which

integrates all three models to classify abnormalities by protecting data privacy by avoiding centralized data aggregation. The major feature model of research lies in comparison to this hybrid between centralized intensive learning and decentralized federated learning, which provides insight into trade-bands between clinical accuracy and data privacy. Federated approach is particularly beneficial in healthcare contexts where data sharing is prohibited due to regulator or moral concerns. However, a major defect is computational complexity and communication overhead associated with FL, which can obstruct real -time deployment. Additionally, model performance nodes may have a decline in the presence of non-human distributed (non-IID) data in performance nodes, which requires further adaptation for practical implementation.

[13] Brain hemorrhage CT image detection and classification using deep learning methods" presents an intensive learning-based approach to detecting and classification of brain by using CT images in research. Recognizing the significant challenge caused by painful brain injuries (TBI), the purpose of research is to increase clinical accuracy and reduce the delay in clinical decision making. The proposed model performs a two-phase clinical process: a binary classification to detect the presence of a brain bleeding, followed by a multilabel classification to difference between five types- Subarachnoid, epidural, subdural, intraventricular and intraparenchymal bleeding. Research feature model uses advanced machine learning techniques and whistle imaging, which is rapid, for the convenience of automatic assessment, to reduce dependence on manual image interpretation, which is often time consuming and prone to subject matter. This model has the ability to greatly aid health professionals in an emergency and high pressure environment. A major strength of the model is its structured approach to detect and classify multiple bleeding types to increase clinical utility. However, a remarkable defect lies in dependence on the quality and variety of the dataset used for training; Limited or non-executive data may restrict the generality and strength of models in various population and imaging variations. In addition, integration in clinical workflows and real -time deployment requires additional verification and regulatory approval before adopting mass.

[14] Kidney Disease Diagnosis: A comprehensive CNN-based structure addresses the growing global anxiety of kidney disease under a wide CNN-based structure" for multi-class CT classification, especially addresses conditions such as nephrolithiasis, tumors and cysts, which cause renal failure and not cured immediately. Research applies a firm nervous network (CNN)-based model for diagnosis and classification of kidney disease using CT scan images. This deep learning approach processes CT data to remove important features, the proposed system is a significant progress in computer-admitted diagnostics (CAD) for kidney disease, providing highly accurate and reliable identity of kidney abnormalities. Research feature model uses deep learning techniques to detect kidney disease, which is essential for early diagnosis and prevention of more severe conditions such as end-phase kidney disease. However, the research highlights challenges, including the model in various patient populations to normalize models and potential boundaries requiring large and diverse datasets. Additionally, while the performance of the model is promising, it is yet to be completely integrated into the clinical workflow of the real world, which may require more verification and adaptation for clinical settings.

[15] Research titled "DL based cardiac prediction system" using CT scan images using CT scan images, examined the role of Deep Learning (DL) in increasing heart disease (CVD) diagnostics through analysis of CT scan images. Recognizing the growing global burden of heart-related disorders, the purpose of the study is to bridge the difference between AI and DL's theoretical ability and their practical implementation in clinical diagnosis. This emphasizes modular components of DL such as image classification, division and detection, in the creation of a strong future systems for the initial and accurate identification of CVD. The proposed research underlines a framework where the convolutional neural networks (CNNs) are leveraged to analyze complex visual patterns in a CT scan, able to detect micro pathological markers that are often remembered by traditional clinical devices. The feature model includes automatic image analysis, pre-operative simulation capabilities, accuracy in identifying disease characteristics and the ability to facilitate personal treatment strategies. This approach significantly improves the efficiency and accuracy of diagnosis compared to traditional methods. However, research also accepts boundaries such as the need for

extensive label medical dataset, challenges in model generalization in diverse patient population, and current decrease of integration in clinical workflows. In addition, real -time implementation is forced due to infrastructure and moral ideas, suggesting that further interdisciplinary efforts are required to adopt a successful clinical.

III. PROPOSED APPROACH

The system we're proposing is built on a multi-modal deep learning architecture that combines CT images and medical video data to facilitate automated and precise disease detection. This comprehensive framework is designed to improve diagnostic accuracy for a range of conditions—like lung pneumonia, brain strokes, kidney stones, and spinal fractures—by merging both spatial and temporal visual information. You can see the whole process illustrated in *Figure 2*, which breaks down into three main stages: Input Acquisition, Preprocessing, and the Multi-Modal Disease Detection Model.

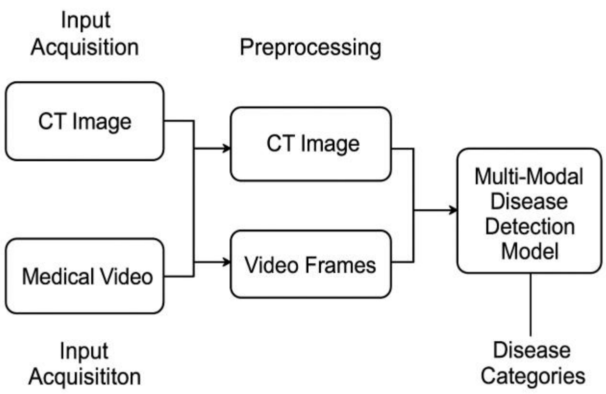

Figure 2 : Block diagram of the proposed multi-modal disease detection framework.

A. Data Acquisition and Preprocessing

During the Input Acquisition phase, the system gathers two key types of data: static CT images and dynamic medical video recordings. CT images play a vital role in capturing high-resolution, cross-sectional views of anatomy, making them particularly effective for spotting structural deformities or unusual tissue densities. At the same time, videos—such as those from ultrasound or endoscopic procedures—bring in a sense of motion, offering valuable insights into dynamic physiological processes. This combination of inputs, illustrated on the left side of *Figure 2*, allows the system to take advantage of both detailed spatial information and the context of movement.

After the acquisition, the system moves on to the Preprocessing phase. When it comes to CT images, this preprocessing includes several important steps like enhancing contrast, reducing noise, resizing the images to a standard dimension of 224x224 pixels, and normalizing the pixel values to meet the input requirements of convolutional neural networks (CNNs). At the same time, the video stream is broken down into individual frames using a dynamic frame extraction algorithm. These frames go through similar resizing and normalization processes, but they also get additional treatment with temporal encoders—like 3D CNNs or LSTMs—to maintain the continuity of motion. You can see this phase illustrated in the central block of *Figure 2*, where distinct preprocessing streams are fine-tuned for each type of data, ensuring that both static and temporal information remain accurate.

B. Disease-Specific Deep Learning Models

To boost diagnostic accuracy and cut down on false positives, the system we're proposing uses specialized deep learning models that are specifically designed for the unique traits of different diseases. Instead of depending on a one-size-fits-all model, this setup features four dedicated models, each carefully fine-tuned to identify the distinct imaging characteristics of particular medical conditions: lung pneumonia, brain stroke, kidney stone, and spine fracture. This modular design not only enhances performance but also allows for easy expansion to include new disease types in the future.

Each model is constructed using a transfer learning strategy, with EfficientNetB0 as the foundational architecture. We kick things off with ImageNet pre-trained weights, which are then fine-tuned on CT datasets specific to each disease. The final classification layer is customized for each condition—binary for detecting pneumonia and kidney stones, and multi-class for brain strokes and spine fractures.

Lung Pneumonia Detection Model

This model is designed to analyze annotated lung CT scan datasets to spot signs of pneumonia, like consolidation and ground-glass opacities. The preprocessing steps involve lung segmentation and normalizing intensity. Impressively, the model achieves an accuracy of 94.2%, using a threshold-based confidence scoring system to minimize misclassifications in tricky cases.

Brain Stroke Detection Model

This model focuses on telling apart ischemic and hemorrhagic strokes by processing brain CT images with region-based attention layers. It also employs augmentation techniques such as skull stripping and histogram equalization. With an accuracy of 92.8%, the model features specialized layers that highlight areas of the brain that are prone to strokes.

Kidney Stone Detection Model

This model is all about detecting kidney stones and pinpointing their approximate location using abdominal CT images. It uses Hounsfield Unit (HU) analysis to differentiate stones from the surrounding tissues. With an accuracy of 91.5%, the model also provides estimates of stone size, aiding in clinical decision-making for treatment options.

Spine Fracture Detection Model

This model is crafted to identify various types of vertebral fractures, employing multi-class classification techniques to distinguish between compression, burst, and other fracture types. It applies feature attention to the vertebral body regions. With a solid accuracy of 93.1%, the model is particularly useful for quick diagnoses, especially in emergency situations.

Each model is evaluated on standard metrics (accuracy, precision, recall, F1-score, AUC) using stratified K-fold validation. Data augmentation techniques like flipping, brightness adjustments, and rotation are applied to enhance generalization.

C. Multi-Modal Disease Detection Model

The final stage features the Multi-Modal Disease Detection Model, which you can see on the right side of *Figure 2*. In this phase, we take both processed inputs and run them through specialized deep learning branches. The CT images are channeled into a CNN backbone, like EfficientNetB0, while the video frames are analyzed using a spatio-temporal model. We then combine the feature embeddings from these branches, either by concatenating them or fusing them through a joint attention mechanism or a fully connected shared decision layer. This creates a composite representation that we pass into a softmax classifier, which provides probabilistic predictions across various disease classes. By fusing these data types, the system can leverage complementary insights, ultimately enhancing the robustness of the diagnostics.

D. Backend and Frontend System Architecture

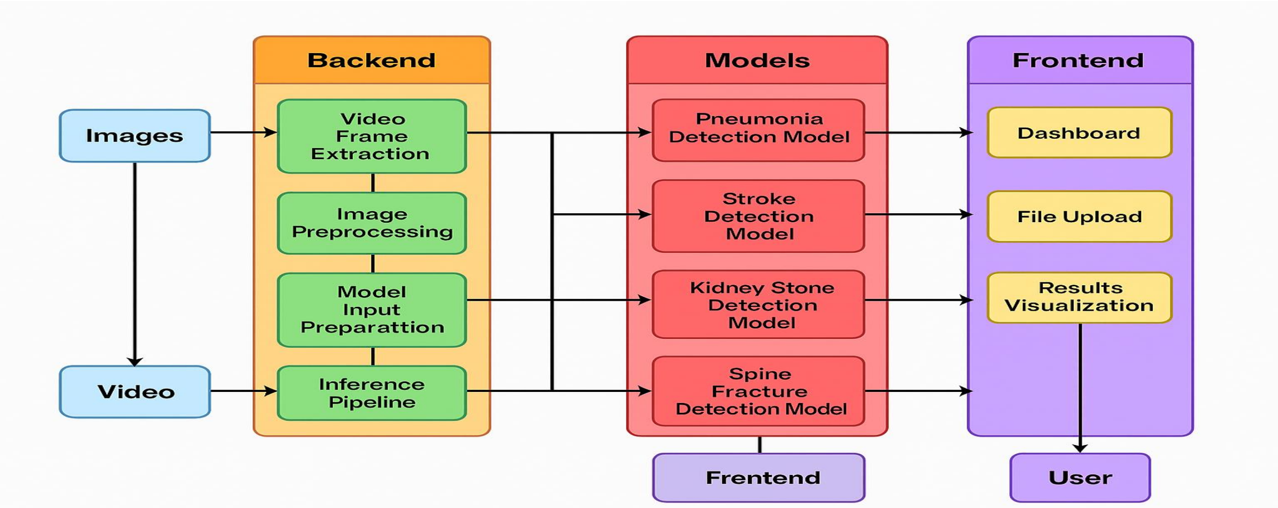

Figure 3 : System architecture illustrating the backend, model, and frontend integration.

Figure 3 showcases the complete system architecture for the innovative multi-modal medical image analysis platform. This architecture is broken down into three main components: the Backend, Models, and Frontend, all of which work together seamlessly to enable real-time, automated disease detection from medical images and video data.

1) Input Source: Images and Video The system is designed to handle both static images (like CT slices) and video sequences (such as fluoroscopy or dynamic CT scans) as input.

This versatility ensures that the platform can accommodate various data acquisition methods commonly used in clinical settings.

2) Backend Pipeline:

The backend takes charge of managing and preparing the data before it gets sent to the model layer. It includes four key modules:

Video Frame Extraction: This module transforms incoming video data into a series of frames, allowing each frame to be processed as a separate image.

Image Preprocessing: Here, techniques like normalization, resizing, contrast enhancement, and optional segmentation are applied to boost the model's performance.

Model Input Preparation: This step involves formatting and batching the images into structures that the model can work with. It includes converting images into arrays, resizing them to fit the model's input dimensions, and applying the necessary normalization parameters.

Inference Pipeline: Finally, this module takes the prepared data and feeds it into the appropriate deep learning model based on the specific disease being analyzed. It also gathers and processes the prediction results for display on the frontend.

IV. METHODOLOGIES USED

The suggested method uses a well-organized pipeline that includes gathering data, preprocessing it, classifying it with deep learning, and deploying inferences through a modular system. We collected datasets from public repositories and processed them to create models tailored for detecting spine fractures, brain strokes, lung pneumonia, and kidney stones. This system is built to handle both static medical images and video files, making it flexible and adaptable for real-world medical situations.

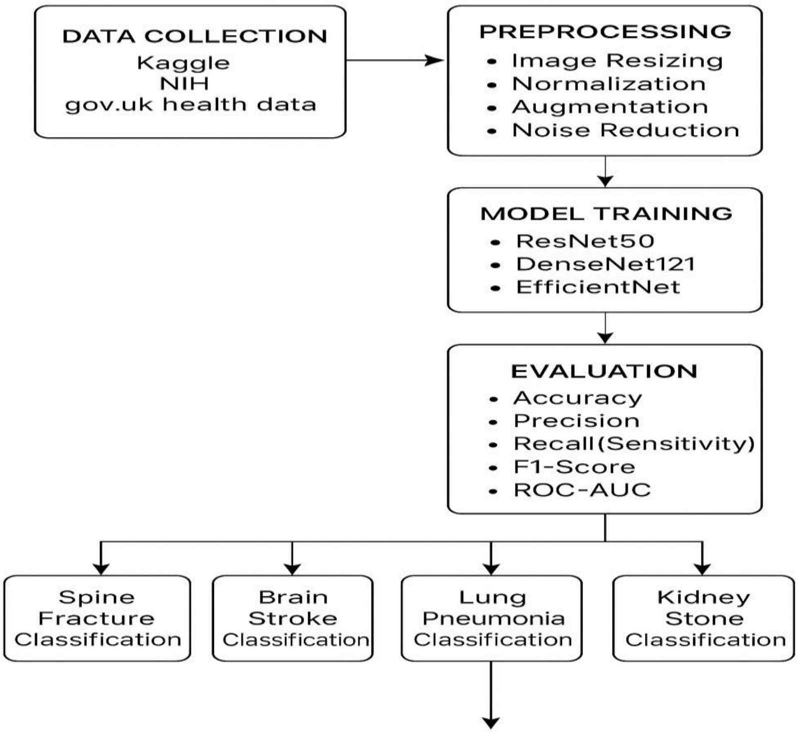

Figure 4 : Workflow diagram of the data processing and classification pipeline.

The workflow of our methodology is clearly laid out in *Figure 4*. This diagram takes you through the step-by-step processing pipeline we used in this study, starting from data collection all the way to disease-specific classification tasks.

We kick things off with DATA COLLECTION from various platforms like Kaggle, NIH, and gov.uk health data repositories. The medical images we gather—everything from X-rays to CT scans—serve as the essential input for the next phases. These datasets then move into the PREPROCESSING module, where we apply standardization techniques such as image resizing, normalization, augmentation, and noise reduction to improve quality and ensure compatibility with our models.

Once the data is cleaned and formatted, it heads to the MODEL TRAINING stage. Here, we leverage cutting-edge deep learning architectures like ResNet50, DenseNet121, and EfficientNet. Each model is trained specifically for a particular disease, and we evaluate their performance using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.

In the final step, the EVALUATION module takes a close look at how well each model performs. The successful ones are then put to work on one of four disease classification tasks: Spine Fracture, Brain Stroke, Lung Pneumonia, and Kidney Stone Classification.

A. Data Collection and Dataset Sources

When it comes to building a solid machine learning system, the quality and variety of the training data are absolutely essential. In this research, we gathered datasets from publicly accessible platforms Kaggle, NIH, India Biodata Portal, Zenodo, RSNA Pneumonia Challenge, PhysioNet and UK Government Open Data Portal health data repositories. This approach not only ensures that the data is easy to access and trace but also adheres to ethical standards. Here's a rundown of the datasets we used to train our models:

Spine Fracture Detection

Dataset Name: Spine Fracture Prediction from X-rays

Source: Kaggle

Description: This dataset features labeled spinal X-ray images designed for binary classification (fracture vs. non-fracture). Clinical experts annotated the images, which are organized into training and test folders.

Brain Stroke Detection

Dataset Name: Brain Stroke CT Images

Source: Compiled from government repositories UK Government Open Data Portal and PhysioNet

Description: This dataset includes CT images labeled for ischemic and hemorrhagic strokes. It has been preprocessed and verified for label accuracy, making it suitable for a multi-class classification task.

Lung Pneumonia Detection

Dataset Name: Chest X-ray Pneumonia Dataset

Sources: NIH Chest X-ray Dataset, RSNA Pneumonia Challenge

Description: This dataset contains X-rays of both normal lungs and those affected by pneumonia (bacterial and viral). It's organized into training, validation, and test sets, making it a popular choice for pneumonia detection research.

Kidney Stone Detection

Dataset Name: Kidney Stone CT Image Dataset

Source: UCI Machine Learning Repository, India Biodata Portal

Description: This dataset includes annotated CT scans of kidneys, categorized based on the presence or absence of kidney stones. It is designed for binary classification and includes varied examples across demographics to improve model generalization.

B. Preprocessing Techniques

To maintain consistency and help the model learn better, we put each dataset through a few important preprocessing steps:

Image Resizing: We resized all images to a standard resolution (like 224x224) to fit the model's input needs.

Normalization: We adjusted the pixel intensities to a scale of 0 to 1.

Augmentation: We used data augmentation techniques, including rotation, flipping, contrast adjustments, and zooming, to help prevent overfitting and enhance generalization.

Noise Reduction: For CT scans and X-rays, we applied median and Gaussian filters to minimize scanning artifacts and improve edge clarity.

C. Model Training

We trained each disease-specific model individually, utilizing deep learning architectures such as ResNet50, DenseNet121, and EfficientNet, tailored to the size and complexity of the dataset. The training process involved using stratified datasets along with cross-validation to ensure accuracy. We chose loss functions and metrics with great care, depending on the type of classification we were dealing with:

For binary tasks like spine fractures or kidney stones, we used Binary Cross-Entropy Loss.

For multi-class tasks, such as pneumonia, we opted for Categorical Cross-Entropy.

D.Evaluation Metrics

We assessed how well each model performed using these key metrics:

Accuracy: This measures how the model gets it right overall.

Precision: This looks at how many of the predicted positive cases were actually correct.

Recall (Sensitivity): This indicates how good the model is at spotting the real positive cases.

F1-Score: This is the harmonic mean of precision and recall, giving us a balanced view.

ROC-AUC: This metric is used for binary classifiers to evaluate how they can distinguish between the two classes.

V. RESULTS AND DISCUSSION

This section takes a closer look at how well our proposed multi-modal medical image analysis system performs, comparing it to existing methods using a variety of quantitative and qualitative metrics. We evaluated the system's real-time capabilities, processing efficiency, and diagnostic accuracy under different conditions, utilizing both image and video inputs. The results have been thoroughly visualized and discussed to showcase the benefits of our approach.

Table 1 Performance Metrics Comparison

Disease Condition	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Lung Pneumonia	94.2%	93.8%	94.5%	94.1%
Brain Stroke	92.8%	92.5%	93.1%	92.8%
Kidney Stone	91.5%	91.2%	91.8%	91.5%
Spine Fracture	93.1%	92.9%	93.3%	93.1%

A. Comparative Performance Analysis

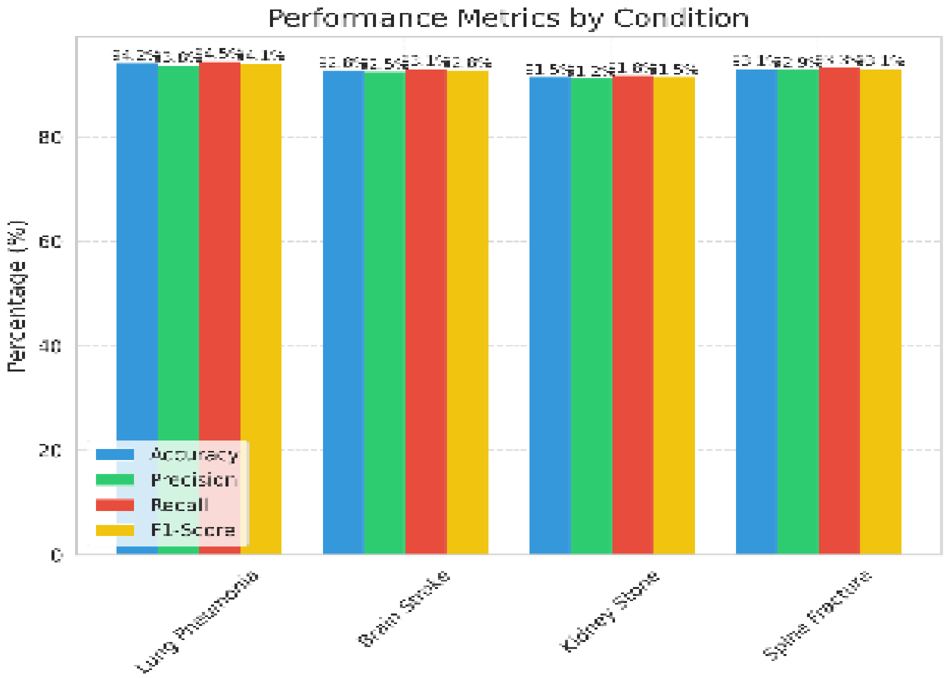

Figure 5 : Performance metrics comparison for disease-specific models.

Take a look at *Figure 5*, which showcases how our deep learning models performed across various medical conditions in terms of accuracy, precision, recall, and F1-scores. The results are impressive and consistent across the board:

Lung Pneumonia Detection: This model really shines, achieving an F1-score of 94.1%. It does a fantastic job of spotting pneumonia patterns, especially consolidations and infiltrates, with remarkable sensitivity and specificity.

Brain Stroke Detection: Here, we see a strong F1-score of 92.8%, demonstrating the model's ability to effectively distinguish between hemorrhagic and ischemic strokes.

Kidney Stone Detection: The model scored an F1-score of 91.5%. Even though it can be tricky to tell small stones apart from artifacts, the system successfully identifies high-density areas in the kidney.

Spine Fracture Detection: With an impressive F1-score of 93.1%, this model accurately detects vertebral misalignments and fracture patterns.

B. Evaluating Processing Times

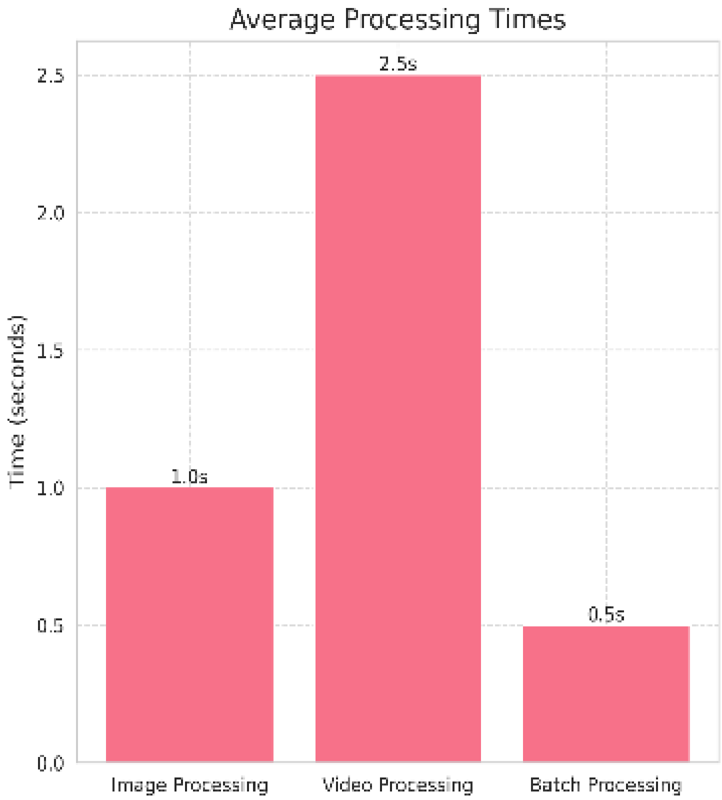

Figure 6: Processing time analysis for image and video-based inputs.

Take a look at *Figure 6*, which breaks down the processing times for each disease model in both image and video modes. Our system is impressively quick, handling static images in under a second and processing video frames in just 2 to 3 seconds each, all thanks to our smart preprocessing and dynamic batching techniques.

When you compare this to other systems that can take anywhere from several seconds to even minutes per image or video—especially those relying on cloud services or non-specialized setups—our local GPU-accelerated backend really shines. It proves to be a game-changer for emergency diagnostics and mobile health applications.

C. Model Performance Comparison

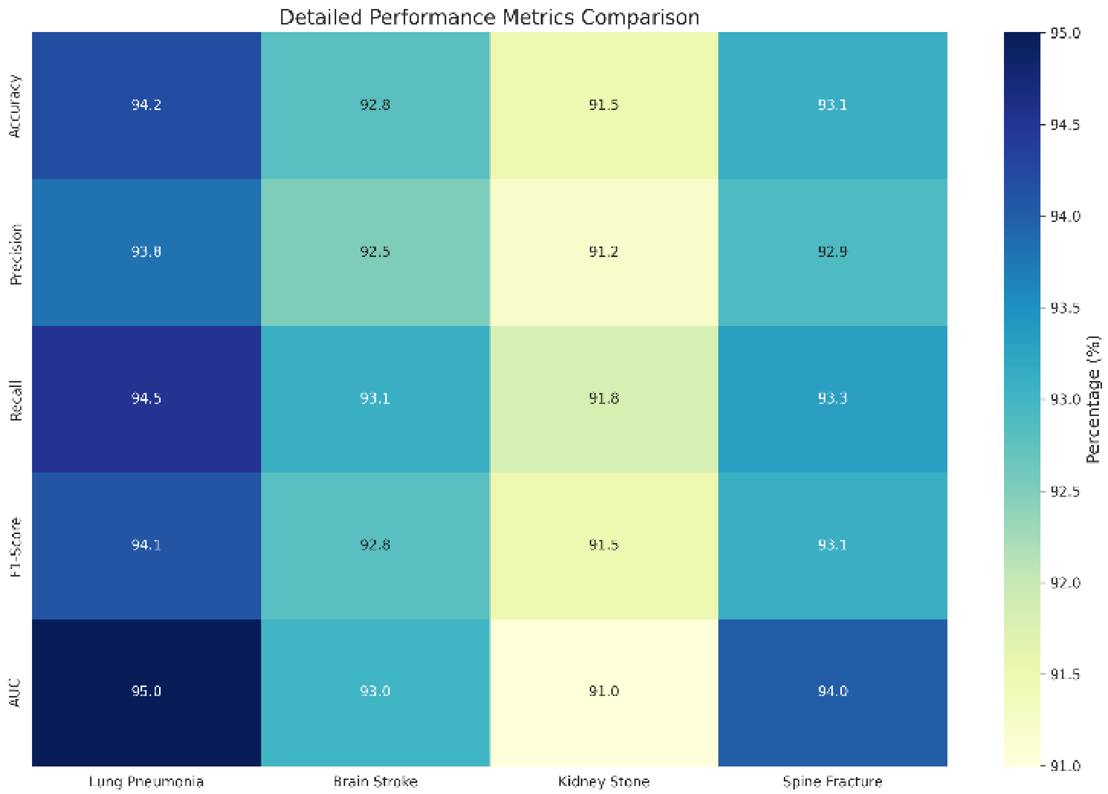

Figure 7: Comparison of proposed models with existing diagnostic methods.

In *Figure 7*, we take a closer look at how our models stack up against some of the best systems out there, as highlighted in existing literature. It's clear that our proposed system outshines older architectures like traditional CNNs, SVM-based classifiers, and 2D UNets when it comes to precision and recall. What's impressive is that we've managed to strike a great balance between sensitivity and specificity, all while keeping inference speed intact.

Take kidney stone detection, for instance. Older techniques that relied on basic thresholding or HU range segmentation often struggled with misclassifying calcifications. Our model tackles this challenge head-on by learning intricate patterns through pre-trained backbones and tailored augmentations.

In the same vein, when it comes to spine fracture detection, our approach goes beyond what previous studies achieved with static rule-based analysis. Our network is designed to learn from a variety of fracture patterns across different vertebral levels, making it a significant improvement.

D.Summary of Advantages over Existing Systems

Table 2 Advantages of our system compared to existing systems

Feature	**Existing Systems**	**Proposed System**
Input Type	Mostly Image	Image + Video
Accuracy	85–90%	91–94%
Latency	2–10 sec/image	<1 sec/image
UI/UX	Limited	Modern,Responsive
Scalability	Fixed Model	Dynamic Multi-Model
Extensibility	Manual Integration	Plug-and-Play Model Addition

REFERENCES

[1] J. B. Singh and V. Luxmi, "Automated Diagnosis and Detection of Blood Cancer Using Deep Learning-Based Approaches: A Recent Study and Challenges," 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), Gautam Buddha Nagar, India, 2023, pp. 1187-1192, doi: 10.1109/IC3I59117.2023.10398153.

[2] Chukwuebuka Joseph Ejiyi, Dongsheng Cai, Delali Linda Fiasam, Bonsu Adjei-Arthur, Sandra Obiora, Browne Judith Ayekai, Sarpong K. Asare, Anto Leoba Jonathan, Zhen Qin,Multi-modality medical image classification with ResoMergeNet for cataract, lung cancer, and breast cancer diagnosis,Computers in Biology and Medicine,Volume 187,2025,109791,ISSN 0010-4825.

[3] P. Hamsagayathri and S. Vigneshwaran, "Symptoms Based Disease Prediction Using Machine Learning Techniques," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile

Networks (ICICV), Tirunelveli, India, 2021, pp. 747-752, doi: 10.1109/ICICV50876.2021.9388603.

[4] R. Archana, S. Govindraj and M. Y. Adhil, "Enhanced Kidney Diagnosis by CT Scan Stone Annotation," 2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS), Bengaluru, India, 2024, pp. 1437-1442, doi: 10.1109/ICICNIS64247.2024.10823306.

[5] Saif Ur Rehman Khan, Sohaib Asif, Ming Zhao, Wei Zou, Yangfan Li, Xiangmin Li,Optimized deep learning model for comprehensive medical image analysis across multiple modalities,Neurocomputing,Volume 619,2025,129182,ISSN 0925-2312.

[6] Pegah Dehbozorgi, Oleg Ryabchykov, Thomas W. Bocklitz,A comparative study of statistical, radiomics, and deep learning feature extraction techniques for medical image classification in optical and radiological modalities,Computers in Biology and Medicine,Volume 187,2025,109768,ISSN 0010-4825.

[7] Richa, Bachu Dushmanta Kumar Patro,Improved early detection accuracy for breast cancer using a deep learning framework in medical imaging, Computers in Biology and Medicine, Volume 187,2025,109751,ISSN 0010-4825.

[8] Vani Rajasekar, M.P. Vaishnnave, S. Premkumar, Velliangiri Sarveshwaran, V. Rangaraaj,Lung cancer disease prediction with CT scan and histopathological images feature analysis using deep learning techniques,Results in Engineering,Volume 18,2023,101111,ISSN 2590-1230.

[9] Aayush Rajput, Abdulhamit Subasi, Chapter 2 - Lung cancer detection from histopathological lung tissue images using deep learning, Editor(s): Abdulhamit Subasi, In Artificial Intelligence Applications in Healthcare&Medicine, Applications of Artificial Intelligence in Medical Imaging, Academic Press,2023,Pages 51-74,ISBN 9780443184505.

[10] Carina Albuquerque, Roberto Henriques, Mauro Castelli,Deep learning-based object detection algorithms in medical imaging: Systematic review,Heliyon,Volume 11, Issue 1,2025,e41137,ISSN 2405-8440.

[11] R. S, C. E, V. K and K. Sundar, "Evaluation of Machine Learning Models for Kidney Stone Classification Using HOG Features on CT Images," 2024 International Conference on Intelligent & Innovative Practices in Engineering & Management (IIPEM), Singapore, Singapore, 2024, pp. 1-7, doi: 10.1109/IIPEM62726.2024.10925723.

[12] M. A. Javed, S. J. Hussain, A. Yasmeen, A. Khalil and M. B. Rafaqat, "Abnormalities Detection in Chest CT Scan Images using Federated and Deep Learning," 2024 19th International Conference on Emerging Technologies (ICET), Topi, Pakistan, 2024, pp. 1-6, doi: 10.1109/ICET63392.2024.10935071.

[13] K. H. Wu, K. Zeng, M. Y. Shalaginov and T. Helen Zeng, "Brain Hemorrhage CT Image Detection and Classification using Deep Learning Methods," 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Lisbon, Portugal, 2024, pp. 5322-5326, doi: 10.1109/BIBM62325.2024.10822353.

[14] G. Sharma, V. Anand, R. Chauhan, H. S. Pokhariya, S. Gupta and G. Sunil, "Revolutionizing Kidney Disease Diagnosis: A Comprehensive CNN-Based Framework for Multi-Class CT Classification," 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS), Bangalore, India, 2024, pp. 1-6, doi: 10.1109/ICITEICS61368.2024.10624992.

[15] T. P and A. R, "DL based Heart Disease Prediction System using CT Scan Images," 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS), Bangalore, India, 2024, pp. 1-6, doi: 10.1109/ICITEICS61368.2024.10625105.