Multimodal Disease Prediction: A Hybrid Approach Combining Machine Learning, Deep Learning, YOLO, and SAM for Diverse Disease Detection

1st Sushma S
Computer Science And Engineering
St. Joseph's College of Engineering
Chennai, India
sushmasivakumar07@gmail.com

2nd Trishala Pandi S
Computer Science And Engineering
St. Joseph's College of Engineering
Chennai, India
trishalasaravanan@gmail.com

3rd Brindha G
Computer Science And Engineering
St. Joseph's College of Engineering
Chennai, India
brindhags@gmail.com

***Abstract*— The healthcare industry faces significant challenges due to fragmented data systems, which impede the ability to obtain a comprehensive view of a patient's health history. Diagnosing multiple conditions accurately is complicated by overlapping symptoms and the presence of comorbidities, necessitating complex care management. Resource constraints, particularly in rural and underdeveloped areas, delay diagnosis and treatment, while variability in clinical expertise leads to inconsistent care quality. Furthermore, many patients are unable to afford modern diagnostic tests due to their excessive costs. This paper introduces an approach that presents a comprehensive healthcare platform, utilizing advanced deep learning algorithms for the early and accurate prediction of multiple diseases. By consolidating data and predictions in a single interface, it provides a holistic view of a patient's health, enhancing diagnostic efficiency and speed. The system's comprehensive disease coverage addresses overlapping symptoms, facilitating better management of comorbidities. This integrated approach not only improves patient outcomes by enabling timely interventions but also offers a cost-effective solution by reducing the need for multiple, expensive diagnostic tests. Our findings indicate that the integrated approach not only improves diagnostic accuracy but also provides a user-friendly interface for patient interaction, thereby contributing significantly to the early detection and management of multiple diseases. Through this analysis, YOLOv10 demonstrated a recall of 0.828. The system achieved a prediction accuracy of 0.9 for obstructive pulmonary disease, 0.98 for brain stroke, 1.0 for diabetes and thyroid disease, 0.84 for heart disease, and 0.99 for liver disease. This system, which makes preventive and individualized medical treatment possible, has the potential to completely transform the way healthcare is provided.**

Keywords—YOLOv10, Segment Anything Model, Multiple Disease prediction, Machine Learning Algorithms.

I. INTRODUCTION

Detecting multiple diseases within individuals presents a significant challenge in modern healthcare due to overlapping symptoms, varied disease pathways, and potential co-existing conditions. Traditional medical practices often focus on diagnosing single diseases in isolation, which may overlook concurrent ailments. However, accurately diagnosing multiple diseases simultaneously is crucial, as it directly impacts treatment efficacy, patient outcomes, and healthcare burden. With 17.9 million deaths a year, or about 31% of all deaths, cardiovascular diseases (CVDs) are the leading cause of mortality worldwide. Cancer follows closely, claiming nearly 10 million lives each year. Respiratory diseases, including COPD [1] and pneumonia, substantially contribute to morbidity and mortality, affecting millions. Additionally, diabetes, impacting approximately 537 million adults, highlights a growing global health emergency due to its rising prevalence. This paper introduces an integrated system aimed at predicting a spectrum of critical conditions, including brain disorders, liver disease, lung disease, thyroid disease, diabetes, heart disease and chronic kidney disease. Recognizing the importance of early detection and diagnosis in potentially fatal disorders, our research employs a range of advanced deep learning algorithms, including Voting Classifier, Stacking Classifier, YOLO(You Only Look Once), SAM(Segment Anything Model) alongside traditional classification algorithms like K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Logistic Regression, and Gaussian Naive Bayes.

This paper is organized into following : Section II reviews the literature on related work. Section III provides a detailed description of the architecture of the YOLOv10 and SAM models. Section IV outlines the implementation process and describes the algorithms selected for this study. Section V presents the results and discusses the findings of the comparative study. Section VI addresses the limitations and challenges in real-world implementation. Section VII discusses the integration of the system into healthcare workflows, particularly in rural and underdeveloped areas. Finally, Section VIII presents the conclusions drawn from the study.

II. LITERATURE REVIEW

The conference papers related to multiple disease prediction [2] that are similar to this study are listed in the above Table I . A brain tumor classification model was created by Hasan Sarker, Mohammad Asif Hasan, and Nayan Roy [3] utilizing ensemble learning with machine learning and Transfer Learning models. With a soft voting classifier to improve performance and MobileNet as the best feature extractor, their model yielded an accuracy of 98.01%. Additionally, a user- friendly web application was developed to help physicians diagnose brain tumors. K. Poorani and M. Karuppasamy [4] developed a KNN algorithm upgrade to enhance recognition and early detection Type 1 Diabetes (T1D), enhancing prediction accuracy. According to studies, early intervention is crucial for personalized treatment strategies for better management of T1D. A comparison analysis of models for machine learning for cardiovascular

TABLE I. COMPARATIVE STUDY

Author(s)	Name of Paper	Results and Analysis
Maria Alex Kuzhippallil, Carolyn Joseph, A . Kannan. [5]	Comparative Analysis of Machine Learning Techniques for Indian Liver Disease Patients	Introduced an extended XGBoost classifier with genetic algorithms for early detection of liver diseases, comparing various models and visualization techniques. Outlier detection using isolation forest enhances accuracy and efficiency in disease prediction.
Reddyvari Venkateswara Reddy, Monisha GS, Ahmed MudassarAli, Palivala YaminiDevi, MadhusudhanMV, Yadala Sucharitha. [6]	Empirical Method for Thyroid Disease Prediction Using a DeepLearning Techniques	Enhanced automatic thyroid nodule recognition in ultrasound images through machine learning strategies achieves 99% accuracy and 97% specificity with the refined random forest algorithm.
Binila Mariyam Boban, Rajesh Kannan Megalingam. [7]	Lung Diseases Classification based on Machine Learning Algorithmsand Performance Evaluation	Integrated and enhanced ML algorithms effectively classify lung diseases using 400 CT scan images, achieving 98% accuracy with MLP and 99.2% with KNN, highlighting their effectiveness in accurate disease classification.
Rishi Singhal, Shailender Gupta, Poonam Singhal [8]	An Ensembling Approach using DL & Non-DL Techniques for Detecting Brain Tumors using MRI Scans	The proposed system detects brain tumors using a hybrid approach combining non-DL features, such as statistical, image-based, and Topological Data Analysis (TDA), with DL techniques. Non-DL features are embedded in MRI scans and classified using a CNN-based model.

disease prediction was carried out by B Anish Fathima, R Vikram, M Sri Vishnu, Sushaanthan R T, and C Venumadhav [9]. The SVM algorithm had the highest efficiency, at 94%, out of all the approaches they studied, including Random Forest, SVM, decision trees, and logistic regression. Krishna Reddy Papana and S Nagakishore Bhavanam [10] developed an Optimized Cascade and ElmanNeural Network (OCENN) for brain stroke classification usingCT images. Their method includes noise removal, lesion segmentation, and feature extraction, achieving 98.60% ac- curacy and outperforming conventional techniques in stroke detection. Existing healthcare systems often focus on a single disease, depend on conventional algorithms and need a large amount of manual input, increasing workload and potential errors. They lack integrated functionalities for comprehensivedisease prediction and personalized drug recommendations. In contrast, the proposed system predicts multiple diseases, including brain tumors, Alzheimer's, liver, kidney, pancreatic, and lung conditions, employing cutting-edge machine learning and deep learning models. It integrates disease prediction and drug recommendations into a single web-based platform with a chatbot with an intuitive user interface [11], removing manual entry and improving early detection and patient care.

III. MODEL ARCHITECTURE

1) YOLOv10: The YOLOv10 model architecture introduces a more complex and efficient backbone network that enhances feature extraction and improves the head design for greater precision and speed. YOLOv10 achieves high detection performance while minimizing computational costs by removing the need for Non-Maximum Suppression (NMS) and incorporating an advanced filtering technique.

NMS traditionally removes redundant bounding boxes, retaining only the most accurate prediction for each object. In YOLOv10, anchor-free techniques and dynamic anchor assignment improve adaptability to various datasets. Optimized loss functions, like focal loss and IoU-based loss, contribute to this performance by balancing localization, classification, and confidence ratings. This performance boost is supported by sophisticated data augmentation, transfer learning, and multi-scale predictions using Feature Pyramid Networks (FPN) and Cross-Stage Partial Networks (CSPNets).

NMS-Free Efficiency: Achieving high efficiency without NMS involves using double labeling and consistent matching measures for steady equalization in training. NMS-free learning typically improves inference efficiency by handling repetitive predictions without additional post-processing. Previous efforts to address repetitive predictions using one-to-one matching methods saw limited success, often at the expense of higher inference costs. YOLOv10's NMS-free training, with dual-label assignments and consistent matching, strikes a balance between efficiency and performance.

Lightweight Classification Head: The YOLOv10 lightweight classification head uses two depth wise separable convolutions followed by a 1×1 convolution, which reduces computational costs. Despite similarities in design, classification and regression heads in YOLO models differ in computational load. For example, in YOLOv8-S, the classification head's FLOPs and parameter count are significantly higher than the regression head's. Analysis shows regression error generally outperforms classification error, allowing classification overhead to be reduced without compromising performance.

Spatial-Channel Decoupled Downsampling: YOLOv10 introduces a spatial-channel decoupled downsampling approach that retains more information while minimizing processing costs by decoupling spatial and channel operations. Traditional YOLO models perform spatial reduction and channel expansion simultaneously using 3x3 filters with stride 2, which increases computation and parameter count. In contrast, YOLOv10 applies pointwise convolution over channels first, followed by depth wise convolution for spatial downsampling, which reduces computational demands, latency, and parameters without losing performance.

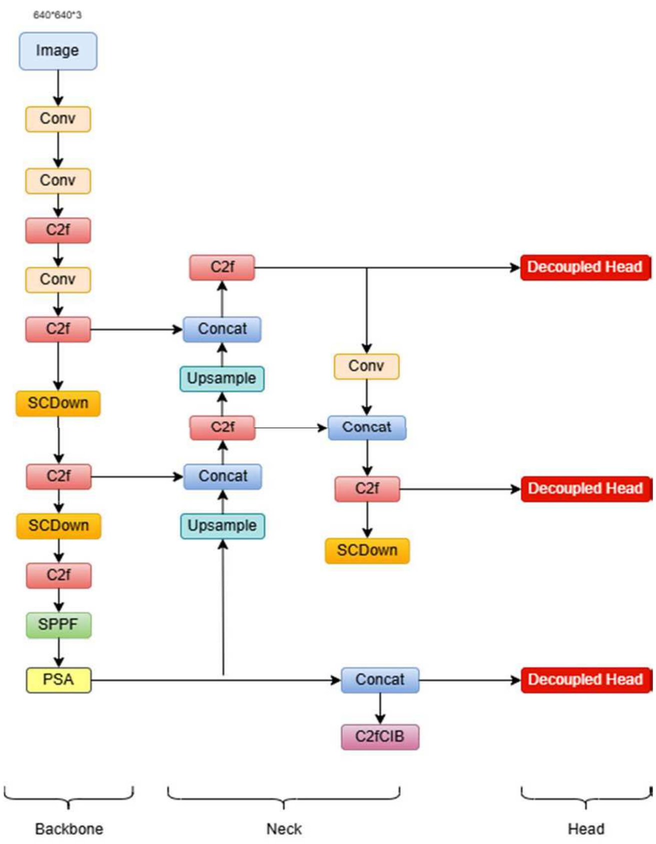

Figure 1. YOLO ARCHITECTURE [12]

2) Segment Anything Model: Image segmentation is the process of breaking up a digital image into many parts, with each region's pixels sharing semantic characteristics. This technique is crucial for analyzing and interpreting images as it groups pixels with similar attributes, making it easier to identify meaningful structures. The Segment Anything Model (SAM) is a powerful tool for precise and adaptable image segmentation, combining several sophisticated components to achieve high accuracy.

Backbone Network: SAM's backbone network can be based on a convolutional neural network (CNN) like ResNet or a vision transformer (ViT) such as the Swin Transformer. ResNet uses deep residual connections, which help train very deep networks by mitigating the vanishing gradient problem and allowing the network to learn complex features. Conversely, vision transformers utilize self-attention mechanisms to capture long-range dependencies and the segmentation quality. contextual information across the entire image, enhancing (i) Prompt Encoder: SAM's prompt encoder lets the segmentation process be guided by user-specified inputs. These cues can be points, boxes, or masks, among other shapes.

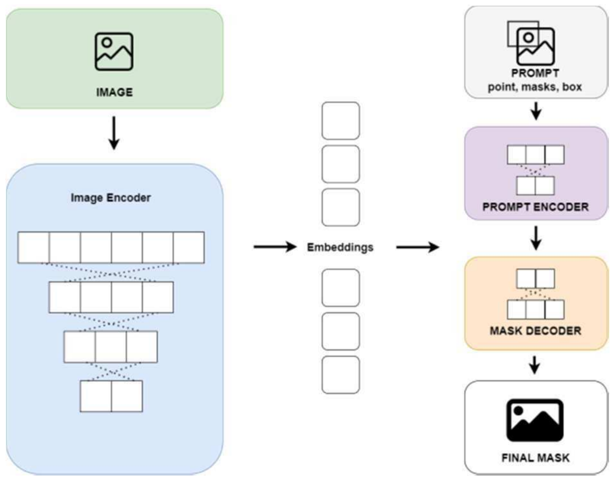

Figure 2. SAM ARCHITECTURE [13]

(ii) Point Encoder: Converts coordinates of specified points into a feature space.
(iii) Box Encoder: Processes bounding box coordinates to guide segmentation.
(iv)Mask Encoder: Integrates binary masks that highlight specific regions or objects. These encoded prompts are combined with image features to precisely direct the segmentation process.

Feature Pyramid Network (FPN) : To handle varying object scales and resolutions, SAM employs a Feature Pyramid Network (FPN). By combining data from different backbone network layers, the FPN captures features across multiple scales. This multi-scale representation is crucial for accurately segmenting objects of different sizes and ensu- res a detailed and robust segmentation output.

Transformer-Based Decoder : The refined features and prompts are processed by SAM's transformer-based decoder, which uses self-attention mechanisms to improve segmentation accuracy. Self-attention captures intricate relationships between prompts and image elements, allowing the model to focus on both simultaneously. Multi-head attention further enhances this by enabling the model to attend to multiple aspects of the image and prompts.

Segmentation Head: The model creates the segmentation masks at the segmentation head, which is the last step. Using convolutional layers, it refines the output to produce high-resolution masks that clearly distinguish the objects within the image. The activation function, typically sigmoid or softmax, is applied to create binary or multi-class segmentation masks.

IV. IMPLEMENTATION

The dataset for the multi-organ disease prediction system is sourced from reputable repositories such as the UCI Machine Learning Repository, Kaggle, and the Google Dataset Search engine, covering diseases across various organs including the brain, kidneys, lungs, liver, heart, pancreas, and throat. This comprehensive dataset encompasses both image and CSV files, ensuring diverse and relevant data collection. Image preprocessing techniques are applied, including resizing to standard dimensions, noise reduction using filters, and normalizationto a standard pixel value range. Gaussian blurring is employed to reduce noise and enhance clarity, aiding in accurate disease detection. The images are cropped to focuson relevant areas and undergo color space conversion and histogram equalization to improve feature extraction and contrast. Edge detection algorithms are utilized to highlightobject boundaries for clearer delineation. The dataset is divided into training and testing sets, with metadata provided in CSV files to support systematic analysis. The model used for tumor detection is trained with YOLOv10, using a learning rate of 0.01, a batch size of 16, and trained over 70 epochs. Machine learning models are trained with optimized hyperparameters to enhance predictive performance. For image-based analysis, preprocessing techniques such as resizing, noise reduction, and normalization areemployed to prepare the data. Feature extraction methods identify significant patterns within the images, aiding in accurate disease prediction. Evaluation metrics like precision, recall, F1 Score, and mean Average Precision (mAP) are utilized to assess model performance.

A. Datasets

The curated datasets includes a diverse collection of text and image data, meticulously preprocessed to enhance analysis and model development. The description of the dataset is given in the TABLE *II*. The table outlines datasets and algorithms utilized for multi-class disease prediction, showcasing diverse approaches to model development. The brain disorder dataset includes 1,101 annotated images, divided into training, validation, and testing sets, featuring MRI and CT scans for tumor detection using YOLOv10. For heart disease, a dataset with 76 attributes, including clinical (e.g., blood pressure, cholesterol) and demographic features, emphasizes predictive modeling with K-Neighbors Classifier. Thyroid disease data, tailored for training Artificial Neural Networks (ANNs), comprises 3,772 training and 3,428 testing instances across hyperthyroidism, hypothyroidism, and euthyroid cases, utilizing hormone levels (e.g., TSH, T3, T4) for Stochastic Gradient Boosting. The obstructive pulmonary disease dataset captures 24 variables, including pulmonary function and quality-of-life metrics, analyzed via a Voting Classifier. Liver disease data, with over 20,000 training samples and features such as bilirubin and alkaline phosphatase levels, is modeled using K-Nearest Neighbors. The diabetes dataset, consisting of 768 instances with numeric attributes like glucose and BMI, employs Gradient Boosting for classification. These datasets, processed through various algorithms, emphasize accuracy and tailored predictive approaches for specific diseases.

TABLE II. DATASET DESCRIPTION

Disease	Dataset Sourced	Algorithms
Brain disorder	The dataset comprises 1,101 images, categorized into 642 for training, 234 for validation, and 225 for testing.	YOLOV10
Heart disease	The database, with 76 attributes, is often analyzed using a subset of 14.	K Neighbors Classifier
Thyroid	The database, suited for training ANNs, contains 3 classes, with 3,772 training instances and 3,428 testing instances.	Stochastic Gradient Boosting
Obstructive Pulmonary	The dataset consists of 101 patients with 24 variables, including characteristics, disease severity, co-morbidities, walking ability, quality of life.	Voting classifier
Liver disease	The Dataset contains 20,000 training data points and approximately 1,000 test data points.	K-Nearest Neigbors
Diabetes	The dataset comprises 768 instances with 8 numeric attributes plus a class attribute.	Gradient Boosting

The text data has undergone extensive preprocessing, including tokenization, stopword removal, and normalization to ensure clean and relevant input for natural language processing tasks. The image data comprises various modalities and has been preprocessed with resizing, normalization, and denoising using techniques like Gaussian blur. This comprehensive preprocessing supports robust model training and evaluation, making the dataset an invaluable resource for developing and testing advanced models in both textual and visual contexts.

V. RESULTS ANS DISCUSSIONS

In this study, the dataset is split into training and testing. During training, the 80% of the data is used to teach the model by exposing it to a range of instances and letting it discover patterns and relationships in the information. Through this procedure, the model's accuracy and performance can be enhanced by modifying its parameters. In contrast, testing makes use of the 20% of data that was not seen by the model during training. For instance, This independent set assesses the model's functionality on new, untested data, providing an objective evaluation of its generalization skills. The accuracy of the models are given in the TABLE *III* and visualized Figure 3.

TABLE III. PREDICTION RESULTS

Disease	Accuracy
Heart Disease	**0.84**
Thyroid	**1.0**
Liver	**0.99**
Diabetes	**1.0**
Obstructive Pulmonary	**0.9**
Brain Stroke	**0.98**

The drug recommendation systems designed to assist healthcare providers in selecting appropriate medications, precautions, exercises, and diets based on specificdiseases. The training dataset for the drug recommendation system is split into 70% for training and 30% for testing. It contains a variety of symptoms and the related target diseases.After a number of models, including Random Forest, SVC, and Gradient Boosting, were tested and found to have an accuracy of 1.0, SVC was chosen because to its effectiveness. The system will recommend suitable medications, precautions, exercises, and diets when a certain disease is selected.

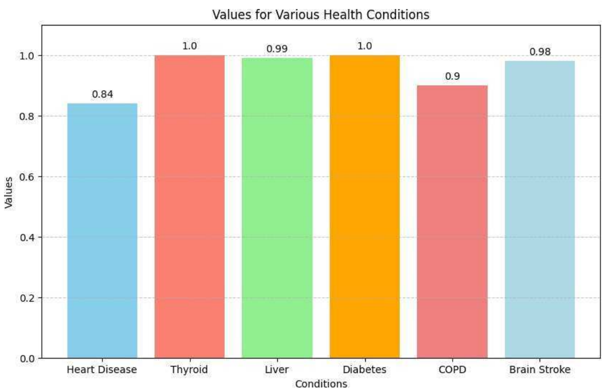

Figure 3. ACCURACY COMPARISON

The YOLO model uses medical images to detect brain tumors, generating a bounding box to classify them as positive or negative. SAM is used to detect small and localized diseases, providing a clear visual depiction of the disease's presence and extent. This technique aids in accurate diagnosis and analysis, ensuring a clear and accurate representation of the disease's extent as seen in brain tumor detection Figure 5 and lung nodule prediction Figure 6. This procedure offers an unambiguous visual depiction of the disease's presence and extent, facilitating accurate diagnosis and analysis.

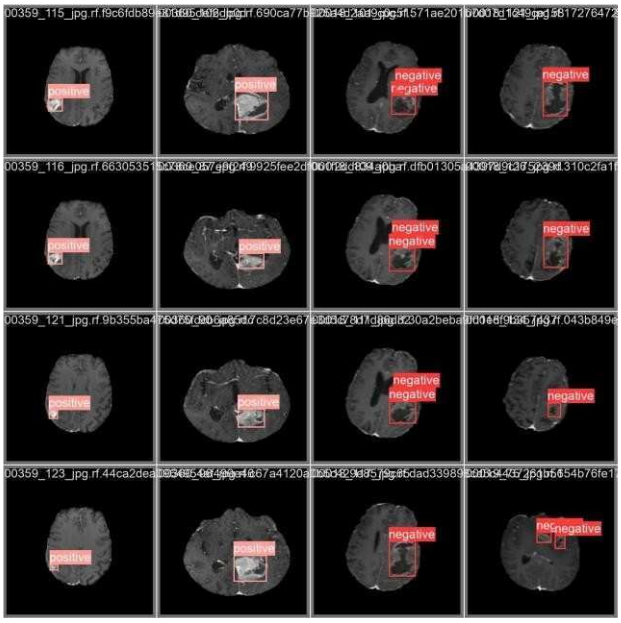

Figure 4. PREDICTED IMAGES

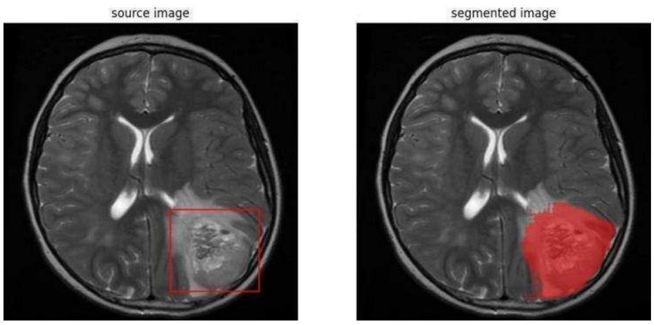

Figure 5. SEGMENTED IMAGES

VI. CHALLENGES IN REAL WORLD IMPLEMENTATION

Implementing the proposed integrated system for multi-disease prediction in real-world settings presents several challenges. The high computational demands of advanced models like YOLO and SAM require significant processing power, which can lead to latency issues, particularly when processing large volumes of high-resolution medical data. These demands necessitate access to high-performance GPUs and cloud computing services, resources that many healthcare facilities, especially those with limited budgets, may find difficult to procure. Furthermore, the complexity of ensemble methods, such as voting and stacking classifiers, increases the computational load, creating a trade-off between achieving high accuracy and maintaining real-time processing speed. This challenge becomes more pronounced as the system scales to predict a larger number of diseases.

VII. RURAL HEALTHCARE INTEGRATION

Integrating the system into rural healthcare systems requires innovative strategies to overcome infrastructure limitations and ensure accessibility. Low-cost devices like Raspberry Pi can perform basic on-site analysis, reducing reliance on expensive high-performance hardware. A hybrid approach can be employed, where simpler tasks are processed locally while complex computations are offloaded to cloud servers

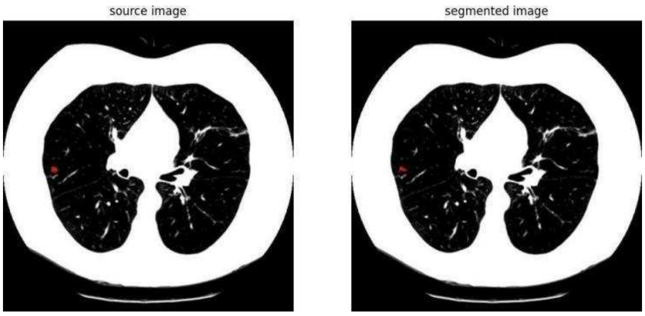

Figure 6. LUNG NODULE PREDICTION

when internet connectivity is available. For areas with intermittent or unreliable connectivity, results can be queued and synchronized once access is restored. Simplified APIs and intuitive user interfaces can facilitate integration with existing healthcare networks, ensuring usability for non-technical healthcare workers. Additionally, training local healthcare providers through hands-on workshops is essential to enable effective usage, building trust and acceptance within the community. By focusing on cost-effective solutions and empowering local personnel, the system can bring advanced healthcare capabilities to underserved regions.

VIII. CONCLUSION

In conclusion, our integrated system combining deep learning and conventional algorithms offers a significant advancement in multi-disease prediction and personalized healthcare. By leveraging sophisticated feature extraction, ensemble learning techniques, and various datasets, this platform has been created to enhance diagnostic precision and supports individualized care. The inclusion of an intuitive chatbot and drug recommendation system provides users with tailored medical advice, treatment recommendations, and lifestyle suggestions, making healthcare more accessible and proactive. This comprehensive approach holds the potential to improve early disease detection, patient health management, and overall treatment outcomes, particularly in underserved regions.

REFERENCES

[1] S. Lee, I. S. Lee and S. Kim, "Predicting Development of Chronic Obstructive Pulmonary Disease and its Risk Factor Analysis," *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC),* 2023.

[2] L. D. Gopisetti, S. K. L. Kummera, S. R. Pattamsetti, S. Kuna, N. Parsi and H. P. Kodali, "Multiple Disease Prediction System using Machine Learning and Streamlit," *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT),* 2023.

[3] M. A. Hasan, H. Sarker and N. Roy, "Brain Tumor Detection Using Feature Extraction and Ensemble Learning with a Smart Web Application," *22023 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD),* 2023.

[4] Poorani, K., Karuppasamy, M, "Analysis of Underlying and Forecasting Factors of Type 1 Diabetes and Prediction of Diabetes Using Machine Learning," *Proceedings of the 2nd International Conference on Cognitive and Intelligent Computing. ICCIC 2022,* 2022.

[5] M. A. Kuzhippallil, C. Joseph and A. Kannan, "Comparative Analysis of Machine Learning Techniques for Indian Liver Disease Patients," *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS),* 2020.

[6] R. V. Reddy, M. G. S, A. M. Ali, P. Yamini Devi, M. M. V and Y. Sucharitha, "Empirical Method for Thyroid Disease Prediction Using a Deep Learning Techniques," *2023 IEEE 8th International Conference for Convergence in Technology (I2CT),* 2023.

[7] B. M. Boban and R. K. Megalingam, "Lung Diseases Classification based on Machine Learning Algorithms and Performance Evaluation," *2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India,* 2020.

[8] R. Singhal, S. Gupta and P. Singhal, "An Ensembling Approach using DL & Non-DL Techniques for Detecting Brain Tumors using MRI Scans," *2022 IEEE International Power and Renewable Energy Conference (IPRECON),* 2022.

[9] B. Anishfathima, R. Vikram, S. R. T, M. Sri Vishnu and C. Venumadhav, "A Comparative Analysis on Classification Models to predict Cardio-vascular disease using Machine Learning Algorithms," *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS),* 2022.

[10] K. R. Papana and S. Nagakishore Bhavanam, "Optimized Cascade and Elman Neural Network for CT Image-Based Brain Stroke Classification," *2024 International Conference on Computing and Data Science (ICCDS), Chennai, India,,* 2024.

[11] A. D, M. M, P. T, T. M, P. S. Sherin A and N. E. R, "Revolutionizing Rural Healthcare in India: AI-Powered Chatbots for Affordable Symptom Analysis and Medical Guidance," *2024 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal,* 2024.

[12] Ding, Ao Wang and Hui Chen and Lihao Liu and Kai Chen and Zijia Lin and Jungong Han and Guiguang, "YOLOv10: Real-Time End-to-End Object Detection," 2024.

[13] Girshick, Alexander Kirillov and Eric Mintun and Nikhila Ravi and Hanzi Mao and Chloe Rolland and Laura Gustafson and Tete Xiao and Spencer Whitehead and Alexander C. Berg and Wan-Yen Lo and Piotr Dollár and Ross, "Segment Anything," *arXiv,* 2023.