International Journal of Web of Multidisciplinary Studies
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
website: https://ijwos.com
Volume 3 Issue 8 , August 2026
E-ISSN : 3049-2424
DOI: https://doi.org/10.71366/ijwos

MediAI - AI Powered Multilingual Clinical Decision Support System (CDSS)

Mrs.Swathi D Mahindrakar[1], Vijay V Honnappanavar[2]
[1] *Assistant Professor G M University Davangere*
[2] *Student 4th Semester MCA, Department of Faculty of Computing and IT, G M University Davangere.*

Article Info

Article History:

Published: 19 August 2026

Publication Issue:
Volume 3, Issue 8
August-2026

Page Number:
413-424

Corresponding Author:
Vijay V Honnappanavar

Abstract:

Patient data almost never shows up in one neat package. A single visit can leave behind structured vitals and symptom checklists, a scanned or handwritten lab slip, a medical image, and a handful of spoken complaints, and most AI health tools out there are only really built to deal with one of these at a time. This paper presents MediAI, a modular clinical decision support system (CDSS) that pulls together machine-learning-based disease-risk prediction, OCR-driven laboratory report interpretation, CNN-based medical image classification, an explainability layer built on feature attribution, a constrained large-language-model (LLM) explanation component, and multilingual text and voice interaction in English, Hindi, and Kannada, all inside one clinician-supervised workflow. Instead of letting a general-purpose LLM reason its way toward a diagnosis, we keep the language model on a short leash here: its only job is to translate the outputs of specialized, independently validated models into plain language, and every AI-generated result has to pass a human review step before anyone treats it as a confirmed clinical record. The paper lays out the proposed architecture, situates it against related work in disease-risk modelling, medical OCR, CNN-based imaging, explainable AI, and multilingual clinical speech recognition, and sets out the experimental protocol needed to validate each module.

Keywords: Clinical Decision Support System, Artificial Intelligence, Machine Learning, Explainable AI, Medical Image Analysis, Optical Character Recognition, Large Language Models, Multilingual Healthcare, Speech Recognition, Human-in-the-Loop

1. INTRODUCTION

Healthcare these days really relies on tools but the information about a single patient is all over the place and it is not even in the same format. When you go to the doctor they take your vitals you fill out a symptom checklist they scan your lab report take an image you tell them what is wrong. Then they send you a follow-up message. Each of these things is usually handled by a different digital tool.

There are models that try to predict if you are at risk for a disease tools that can read lab reports programs that can look at images and assistants that can have conversations with you about your health. These things have all been studied,. They are all separate so they do not really work together. For example a model that checks if you are at risk for blood pressure does not look at your cholesterol levels and a program that looks at chest X-rays cannot explain what it means to someone who does not speak English.

The fact that all these tools are separate is a problem, for two reasons.First, patients and clinicians in linguistically diverse settings, India is a clear example, with dozens of major languages and large rural populations far more comfortable speaking than typing in English, are poorly served by tools that quietly assume English literacy and a single mode of interaction. Second, when prediction, explanation, and image interpretation each live in separate

systems, there's no single point where a clinician can look over the whole picture before it reaches a patient, which makes it harder to hold AI-assisted care to any consistent standard of oversight.

MediAI treats these as facets of one architectural problem rather than as separate products to build one after another. The system pairs a MERN-stack (MongoDB, Express, React, Node.js) web application with a FastAPI-based AI service layer, and its AI components follow one consistent principle throughout: specialized, independently evaluable models, a gradient-boosted or ensemble classifier for disease-risk prediction, an OCR-plus-parsing pipeline for laboratory reports, a CNN scoped to one clearly defined imaging task, do the actual analytical work. An explainability layer such as SHAP shows why a model reached a given output, and a large language model (Gemini) is used only to turn that already-computed, structured evidence into a natural-language explanation in the patient's or clinician's preferred language, delivered as text or, through a speech-to-text/text-to-speech pipeline, as voice. Every AI-generated clinical assessment sits in a "pending review" state until a doctor accepts, edits, or rejects it.

The real question this paper is asking is whether an architecture built this way, unified, multimodal, multilingual, can offer genuinely useful decision support while still keeping a clinician in the loop, and whether its individual pieces hold up when judged on their own terms rather than by whether the finished application simply runs. Sections II-IV review related work and pin down the specific gap MediAI is meant to fill; Sections V-VI describe the proposed framework and methodology; Sections VII-IX lay out the experimental protocol, present results, and discuss what they mean; Sections X-XIII cover security, limitations, future work, and conclusions.

2. RELATED WORK

Ten studies were chosen to represent the technical threads MediAI draws on: ML-based disease-risk prediction, OCR-based extraction from clinical documents, CNN-based medical image classification, explainable AI in clinical settings, LLM-assisted clinical decision support, and multilingual or voice-based healthcare interaction.

A. Machine-Learning-Based Disease Risk Prediction

Feng and colleagues trained an ensemble boosted-tree model on a stratified cohort of more than 36,000 hospitalized patients to catch healthcare-associated infection before it became clinically obvious, using vital signs, lab values, and demographics as inputs. Their best model reached a cross-validated AUC of 0.88 a full hour before clinical suspicion arose, and stayed above 0.85 across the preceding 48-hour window. Worth noting: the full feature set could be trimmed down to the 36 most routinely measured variables without any meaningful drop in performance, a useful data point for MediAI's Module 1, which is meant to work from the kind of inputs a clinic can realistically gather rather than an exhaustive test panel.

A separate heart-disease prediction study ran the kind of head-to-head comparison MediAI's Module 1 calls for, testing seven classifiers (Logistic Regression, Random Forest, Decision Tree, Naive Bayes, k-Nearest Neighbors, Neural Networks, and Support Vector Machine), with SVM coming out ahead on accuracy in their setup. It also doubles as a cautionary example: reporting accuracy alone, without recall, F1, or ROC-AUC, is exactly the kind of gap MediAI's evaluation plan is built to avoid, particularly given how misleading accuracy can be on imbalanced clinical outcomes.

B. OCR-Based Extraction from Clinical Documents

Xue and colleagues built a two-stage OCR pipeline, detection followed by recognition, aimed specifically at photographed or scanned laboratory reports, motivated by the everyday problem that complete records are often unavailable at the point of treatment because of gaps in EHR interoperability. Their patch-based detection strategy reportedly achieved 99.5% recall, and the paper is a useful precedent for treating OCR as its own independently evaluated module instead of folding it into a vague "AI reads your report" claim.

Closer still to MediAI's Module 2 is the work of Ma and colleagues, who paired OCR with a named-entity-recognition layer to pull test name, result, unit, and reference range out of paper-based lab reports as four distinct entity types. They point out explicitly that lab report layouts are diverse and non-standard, mixing text, numbers, reference ranges, and units in ways that make naive OCR-to-database conversion unreliable without further extraction logic on top,

which is exactly why MediAI reports "reference range unavailable" rather than guessing at a universal normal range when extraction is uncertain.

C. CNN-Based Medical Image Classification

One study benchmarked several pretrained CNNs on chest X-ray images before fine-tuning the strongest performer into a customized MobileNetV2 variant for multiclass lung-lesion classification, essentially the compare-then-refine approach recommended for MediAI's Module 3. Importantly, the paper scopes its claims to chest X-ray lesion classes only rather than presenting itself as a general-purpose imaging tool, a discipline MediAI's own paper is advised to follow.

A second study trained eleven CNN architectures, including VGG16/19, ResNet50/101, InceptionV3, DenseNet, EfficientNetB7, and MobileNetV2, on 3,616 COVID-19 and 10,192 healthy chest X-rays, then modified the best-performing one further. The resulting modified MobileNetV2 model reached 98% accuracy, the highest of all eleven models tested, ahead of a pretrained MobileNetV2 baseline at 97%. That comparative-then-refine design, together with its reporting of model compilation and training time, is a reasonable template for how MediAI should evaluate its own CNN work and report it in a reproducible way.

D. Explainable AI in Clinical Decision Support

A systematic review by Alkhanbouli and colleagues surveys how techniques such as SHAP and LIME have been applied across disease-prediction studies, pulling together findings on where XAI genuinely builds clinician trust and where it risks oversimplifying what a model is actually doing. This supports MediAI's decision to treat prediction and explanation as two separate, sequential steps rather than asking one model to do both jobs at once.

A more architecturally specific paper proposes a two-stage pipeline where SHAP values are computed first and then used to steer an LLM toward a natural-language explanation, on the reasoning that SHAP identifies which features mattered but is not very readable on its own, while LLMs write readable text but tend toward vague or unfaithful summaries when they are not grounded in something concrete. This is essentially the same separation MediAI proposes for its ML → SHAP → Gemini explanation chain, which suggests the design is a recognized pattern rather than something invented ad hoc.

E. LLMs in Clinical Workflows

A review of LLM-powered clinical decision support spanning oncology, chronic disease management, and digestive disease workflows keeps arriving at the same conclusion: outputs still need expert verification, since LLMs can surface useful information on cancer screening and management but still require a clinician to check it before it becomes a source of misinformation. That pattern repeats across every specialty the review examined. This framing, augmenting clinical judgment rather than replacing it, lines up directly with MediAI's human-in-the-loop review workflow, and supports positioning Gemini as an explanation layer rather than a diagnostic engine.

F. Multilingual and Voice-Based Healthcare Interaction

The paper most directly relevant to MediAI's voice module evaluates ASR performance specifically on multilingual doctor-patient interviews conducted in Kannada, Hindi, and Indian English, using word error rate, character error rate, and substitution-insertion-deletion analysis, while also checking for fairness across speaker gender and role. It compares commercial and open models, including Sarvam's Saarika and Gemini-based speech recognition, in exactly the language set MediAI targets, and effectively hands over a ready-made evaluation methodology (WER/CER by language, fairness by speaker role) that MediAI's own voice-pipeline evaluation (RQ5) can adopt directly.

3. RESEARCH GAP

Taken together, the literature reviewed above shows steady progress within each individual sub-problem, disease-risk prediction, laboratory OCR, medical-image classification, explainable AI, LLM-assisted explanation, and multilingual clinical speech recognition, but very little integration across them inside one clinically supervised system. The disease-

risk and CNN-imaging studies (Sections II-A and II-C) evaluate their models thoroughly but stop short of document understanding, multilingual access, or explanation generation. The SHAP-plus-LLM explanation pipeline in Section II-D is architecturally close to MediAI but is not embedded in a broader system that also handles OCR, imaging, or voice. The multilingual clinical ASR study in Section II-F rigorously evaluates speech recognition in exactly MediAI's target languages, but as a standalone audit rather than as part of an end-to-end pipeline feeding into downstream prediction or explanation. No study reviewed here combines structured-data prediction, document OCR, task-specific imaging, XAI-grounded LLM explanation, multilingual voice access, and a formal clinician-in-the-loop review stage within one evaluated architecture.

This is the gap MediAI is positioned to address: a modular, multimodal, multilingual CDSS architecture combining specialized ML prediction, explainable AI, OCR-based laboratory interpretation, task-specific CNN medical-image analysis, LLM-assisted explanation, multilingual speech interaction, and clinician-in-the-loop validation within one healthcare workflow, evaluated module by module rather than only as a working application.

4. OBJECTIVES

- Design a modular multimodal CDSS integrating structured clinical data, laboratory documents, supported medical images, and natural-language interaction.
- Develop ML-based disease-risk prediction using demographic, symptom, history, and vital-sign features.
- Integrate OCR-based laboratory-report extraction with structured abnormal-value identification.
- Implement CNN-based classification for a specifically scoped, supported medical-image task.
- Use an LLM (Gemini) to turn structured AI/model outputs into understandable clinical decision-support explanations, rather than using the LLM as the sole diagnostic model.
- Provide multilingual text and voice interaction in English, Hindi, and Kannada.
- Introduce explainability mechanisms (e.g., SHAP) for predictive models.
- Implement a clinician-review workflow before AI-generated assessments are treated as reviewed clinical information.
- Evaluate individual AI modules quantitatively, rather than judging the project only by whether the deployed website functions.

5. PROPOSED FRAMEWORK

A. High-Level Architecture

MediAI is organized around three user roles, Patient, Doctor, and Administrator, who interact through a React frontend backed by a Node.js/Express API layer with MongoDB persistence, alongside a FastAPI service that hosts the AI components.

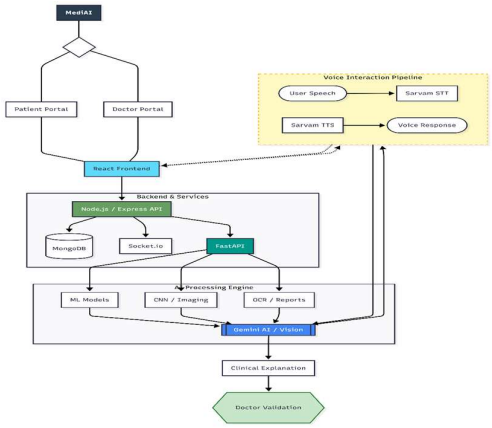

Fig. 1. High-level system architecture of MediAI.

Structured patient variables, age, gender, BMI, blood pressure, blood sugar, heart rate, symptoms, smoking status, family history, and medical history, are cleaned, encoded, and scaled before being passed to a candidate set of models: Logistic Regression, Decision Tree, Random Forest, and XGBoost, selected experimentally rather than assumed in advance (see Sections VI-A and VIII).

B. Explainable AI Layer

Model predictions are paired with a SHAP-based (or equivalent) feature-attribution step, which produces structured evidence, for instance, which features pushed a risk score up or down, before any natural-language explanation is generated. Gemini is prompted only to translate this structured evidence into plain language; it never computes or alters the underlying feature-importance values.

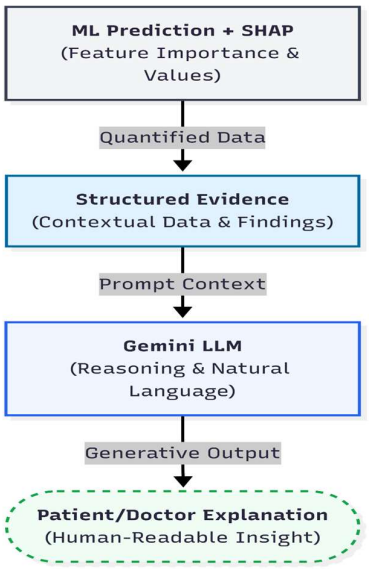

Fig. 2. Explainable AI and Gemini explanation flow.

C. Module 2 - Laboratory Report Analysis

Uploaded PDF or image lab reports (CBC, LFT, KFT/RFT, lipid profile, thyroid profile, glucose, urinalysis) pass through OCR and then a clinical value parser that pulls out test, value, unit, and printed reference range. Extracted values are checked against the printed reference range on the report itself rather than a hard-coded universal range, and if a range cannot be reliably extracted, the system reports “reference range unavailable” instead of guessing.

D. Module 3 - Medical Image Analysis

A CNN, evaluated against alternatives such as MobileNetV2 (Section VI-C), performs classification for one clearly scoped imaging task, chest X-ray screening, for example. Gemini Vision adds a supplementary, non-diagnostic explanation of the CNN's output, but it does not perform the classification itself. The paper deliberately avoids claiming coverage of imaging modalities (MRI, CT, ultrasound, dermatology, retinal imaging) that were not separately trained and evaluated.

E. Module 4 - LLM Explanation Layer (Gemini)

Gemini sits in the system as a constrained natural-language reasoning and explanation layer over the structured outputs of the predictive, OCR, and imaging components, not as an independent diagnostic model. The same “structured evidence → Gemini → explanation” pattern is applied consistently across all three data modalities described in Sections V-B through V-D.

F. Module 5 - Multilingual Voice Interaction

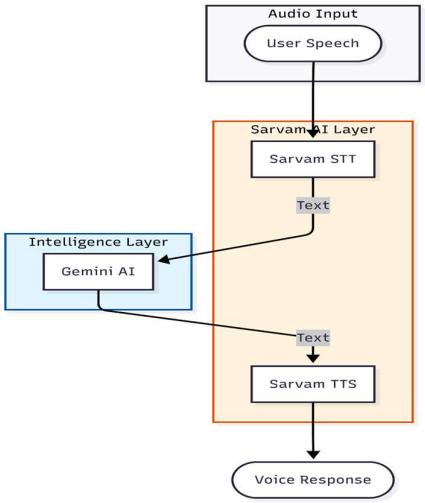

Fig. 3. Multilingual voice interaction pipeline.

The pipeline supports English, Hindi, and Kannada, and the language of the response always matches the language of the input.

G. Doctor Recommendation and Appointment Workflow

Doctor recommendation relies on deterministic, explainable database filtering (specialty, then state, then district, then availability) rather than an LLM, followed by a conventional appointment-booking flow: find doctor, view availability, select slot, book, then view, reschedule, or cancel. This is backed by MongoDB and reflected identically on the doctor's dashboard.

H. Real-Time Communication, Medicine Management, and Health Analytics

Socket.io handles real-time patient-doctor messaging (text, images, documents, read and typing state) with MongoDB persistence and conversation membership enforced on the backend. A medicine-adherence module tracks scheduled doses against taken or missed status. Health analytics only aggregate measurements the patient has actually recorded (BMI, weight, blood pressure, glucose, adherence, risk history); when there is nothing to show, the dashboard displays "no health data recorded yet" rather than a fabricated chart.

I. Human-in-the-Loop Clinical Workflow

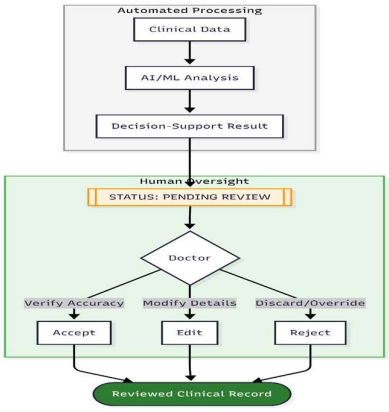

Fig. 4. Human-in-the-loop clinical review workflow.

This workflow is one of the paper's central contributions: no AI-generated result is shown to a patient as a confirmed clinical finding until a clinician has explicitly reviewed it.

J. Database and Security Architecture

MongoDB collections include users, patients, doctors, appointments, medicalReports, labReports, medicalImages, diseasePredictions, medicines, notifications, conversations, messages, and clinicalAIAnalyses. The last of these records model name and version, structured result, Gemini explanation, review status, reviewing doctor, and timestamps, giving every AI-assisted assessment an audit trail.

6. METHODOLOGY

A. Disease Risk Prediction

Candidate models: Logistic Regression, Decision Tree, Random Forest, and XGBoost are trained and compared rather than a single model being chosen by assumption, in keeping with the practice observed in Section II-A.

B. Laboratory Report OCR and Parsing

Pipeline: OCR extraction, followed by clinical value parsing (test, value, unit, reference range), then range comparison against the extracted, not assumed, reference interval, flagged-value output, and finally a Gemini explanation limited to the flagged values.

Ground truth: Manual transcription of each report serves as ground truth for computing Character Error Rate (CER), Word Error Rate (WER), field-extraction accuracy, numeric-value extraction accuracy, and unit-extraction accuracy.

C. Medical Image Analysis

Architecture: A custom CNN and a MobileNetV2-based transfer-learning model are trained and compared, following the compare-then-refine approach used in Section II-C. Preprocessing covers resizing, normalization, and, where applicable, augmentation; training configuration (optimizer, learning rate, batch size, epochs) is reported in full so results can be reproduced.

Scope statement: The prototype is evaluated only on the specific imaging task and classes it was trained on; no claim is made about other modalities (MRI, CT, ultrasound, dermatology, retinal imaging) unless those are separately implemented and evaluated.

D. Explainability and LLM Explanation

SHAP, or an equivalent feature-attribution method suited to the chosen model class, is computed on the disease-risk model's outputs. The resulting structured evidence, not raw model internals, is passed to Gemini with a constrained prompt template designed to keep the explanation consistent with the SHAP output rather than letting it drift into a free-form clinical narrative. A held-out set of Gemini explanations should be checked against their corresponding SHAP evidence for factual consistency (Section VI-F).

E. Multilingual Voice Pipeline

Speech input in English, Hindi, and Kannada is transcribed via Sarvam STT, passed through Gemini for response generation, and returned via Sarvam TTS. Evaluation follows the methodology used in Section II-F: WER/CER by language, end-to-end latency, STT/TTS success rate, and response-language correctness.

F. Ablation and Comparison Design

- Disease prediction: Logistic Regression vs. Decision Tree vs. Random Forest vs. XGBoost.
- Medical imaging: custom CNN vs. MobileNetV2.
- Explanation quality: raw ML output vs. ML + XAI vs. ML + XAI + Gemini explanation (this last comparison needs a structured expert or user evaluation rather than standard classification metrics, see Section VII-D).

7. EXPERIMENTAL SETUP

A. Metrics by Module

TABLE I

Module	Metrics
Disease risk prediction	Accuracy, Precision, Recall, F1, ROC-AUC, Confusion Matrix
Medical image classification	Accuracy, Sensitivity, Specificity, Precision, F1, ROC-AUC
OCR / lab report extraction	CER, WER, Field/Numeric/Unit extraction accuracy
Multilingual voice pipeline	WER by language, STT/TTS success rate, Language correctness, Latency
System performance	API response time, inference latency, OCR time, Gemini latency

B. Human Evaluation of Explanation Quality

A small panel of clinicians and/or patients (target N to be specified) rates a sample of Gemini-generated explanations for factual consistency with the underlying SHAP evidence, clarity, and perceived usefulness, using a structured rubric such as a 1-5 Likert scale per criterion. This addresses RQ4 and is meant as a qualitative complement to the quantitative module metrics above, not a replacement for them.

C. Reproducibility Statement

All random seeds, library versions, and train/test split indices used to produce the results in Section VIII should be recorded and made available, for example, through a supplementary repository, in keeping with the reproducibility emphasis noted in Section II-C.

8. SYSTEM IMPLEMENTATION AND RESULTS

Formal quantitative benchmarking of each AI module, following the metrics and protocol defined in Section VII, is still in progress at the time of writing. Rather than publish placeholder numbers, this section documents the working implementation itself: the interfaces through which the Patient, Doctor, and Administrator roles actually interact with the system described in Sections V and VI. Quantitative results (Tables I-type accuracy, precision, recall, WER, and latency figures) will be reported in a subsequent revision once model training and evaluation are complete.

A. Landing Page and Public-Facing Interface

The public landing page introduces MediAI's core capabilities, AI-assisted health assessment, laboratory report analysis, medical image interpretation, multilingual support, and doctor connectivity, and includes an explicit disclaimer that the platform provides clinical decision support and educational information rather than a substitute for a qualified healthcare professional.

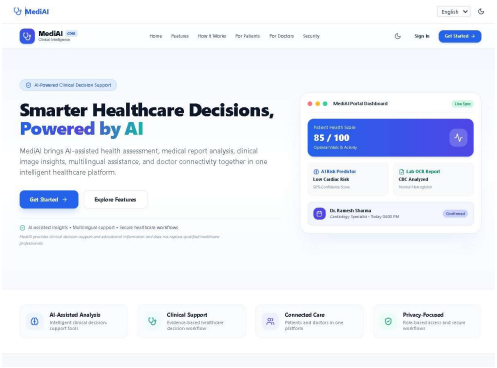

Fig. 5. MediAI landing page and platform overview.

B. Patient Dashboard

The patient-facing dashboard consolidates health score, BMI, blood pressure, and fasting blood sugar into a single view, alongside weekly vitals trends, recent medical reports, upcoming appointments, and the day's medicine schedule. Quick-access actions let a patient jump directly into symptom checking, lab-report upload, image analysis, doctor search, appointment booking, or the AI assistant.

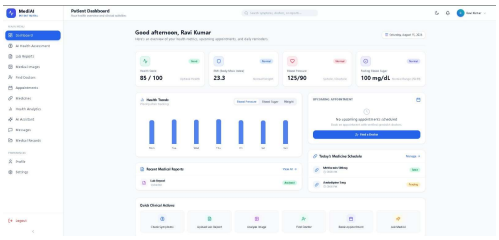

Fig. 6. Patient dashboard showing health metrics, trends, and quick clinical actions.

C. Administrator Dashboard

The administrator view gives platform-level oversight: registered patients, verified doctors, pending doctor-verification requests, hospital network size, and total AI diagnostic analyses performed. A system-health panel confirms the operational status of the Node.js/Express backend, MongoDB instance, FastAPI ML microservice, Gemini Vision integration, and the Sarvam speech pipeline, and an audit stream logs recent platform activity such as doctor-verification approvals.

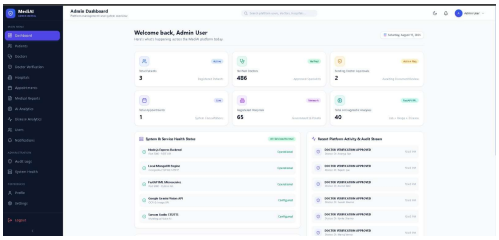

Fig. 7. Administrator dashboard with platform metrics and system-health monitoring.

D. Doctor Dashboard

The doctor-facing dashboard surfaces the day's appointments, active patient count, pending consultations, and completed consultations, together with a weekly appointment-overview chart, the recent-patients list, and a clinical-alerts panel that flags items such as an abnormal fasting-glucose reading or a hypertension follow-up that is coming due. This is the surface through which the human-in-the-loop review described in Section V-I is expected to take place.

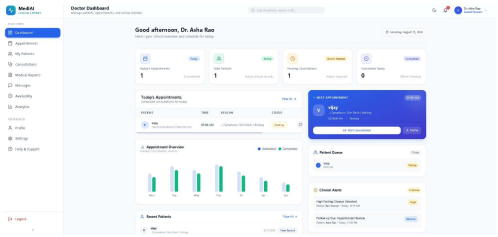

Fig. 8. Doctor dashboard with appointment overview, patient queue, and clinical alerts.

E. Status of Quantitative Evaluation

The interfaces above demonstrate that the end-to-end workflow, from patient-facing data entry through to doctor-facing review, is implemented and functional. They are not a substitute for the module-level metrics defined in Section VII-C: disease-risk classification performance, CNN sensitivity/specificity on the imaging task, OCR character/word error rates, and multilingual voice WER still need to be measured against held-out data before any accuracy claim can be made. That evaluation is treated as the immediate next step for this work rather than as complete.

9. DISCUSSION

The screenshots in Section VIII confirm that MediAI's architecture, three role-based interfaces sitting on top of a shared MERN/FastAPI backend, functions end to end: a patient can reach symptom assessment, lab-report upload, and image analysis from one dashboard; an administrator can see the platform's health and doctor-verification queue; and a doctor can see exactly the patients and alerts that the human-in-the-loop workflow (Section V-I) is designed to route to them. That is a meaningfully different claim from a benchmarking result, however, and the two should not be conflated.

10. SECURITY, PRIVACY, AND CLINICAL SAFETY

- **Authentication:** JWT-based authentication for all user sessions.
- **Authorization:** Role-based access control distinguishing Patient, Doctor, and Administrator permissions.
- **Object-level authorization:** Backend enforcement preventing a doctor from accessing an unrelated patient's record by manipulating a resource identifier in a request.
- **Password protection:** Passwords are hashed, never stored in plaintext.
- **API-key protection:** Gemini and Sarvam API keys remain server-side and are never exposed to the client.
- **File-upload protection:** MIME-type validation, file-size limits, and restriction to permitted document/image categories for lab reports and medical images.
- **Input validation:** Enforced at both the frontend and backend boundary, not the frontend alone.
- **Auditability:** Every AI-generated clinical analysis records which model produced it, its review status, and which clinician reviewed it (Section V-J).
- **Regulatory scope:** This prototype does not claim full compliance with HIPAA, India's Digital Personal Data Protection Act, or equivalent regulatory frameworks unless each applicable requirement has been separately implemented and independently assessed; any such claim should be added only after that assessment is complete.

11. LIMITATIONS

- The disease-risk model is trained and evaluated on the dataset described in Sections VI-A/VII-B, and its performance may not generalize to populations with different demographic or clinical characteristics.
- The medical-image module is scoped to the specific imaging task and classes on which the CNN was trained; it doesn't analyze other imaging modalities unless those are separately implemented and evaluated.

- OCR accuracy depends on image and scan quality; degraded or non-standard lab report formats may reduce field-extraction accuracy, though the actual figure has not yet been measured (Section VIII-E).
- Gemini-generated explanations, despite being grounded in structured SHAP evidence, carry a residual risk of language-model hallucination or oversimplification; the human-evaluation step in Section VII-D is meant to characterize this risk, not eliminate it, once it has been carried out.
- Multilingual voice performance may differ across English, Hindi, and Kannada because of differing training-data availability for each language in the underlying speech models, consistent with disparities reported in related ASR literature (Section II-F).
- The human-in-the-loop workflow depends on consistent clinician engagement; its real-world safety benefit has not yet been validated through a prospective clinical study.
- This work has not undergone clinical deployment or regulatory validation and should be read as a research prototype, not a certified medical device.

12. FUTURE WORK

- External validation on independent, ideally multi-site, patient cohorts to test how well the disease-risk models generalize.
- Expansion of the CNN module to additional, separately validated imaging modalities (e.g., dermatology, retinal imaging) rather than broadening claims without matching evaluation.
- Extension of multilingual support to additional Indian languages, guided by the same WER/CER/fairness evaluation methodology used for English, Hindi, and Kannada.
- Integration with FHIR/EHR standards for real-world interoperability.
- Exploration of federated learning to enable model improvement across institutions without centralizing patient data.
- Deeper explainability methods beyond SHAP (e.g., counterfactual explanations) and formal evaluation of explanation faithfulness.
- Prospective clinical studies assessing whether the human-in-the-loop workflow measurably improves diagnostic accuracy, review time, or patient outcomes compared to standard-of-care workflows.

13. CONCLUSION

This paper set out to present MediAI, a modular clinical decision support system that brings together structured-data disease-risk prediction, OCR-based laboratory report interpretation, task-specific medical-image classification, explainable-AI-grounded LLM explanation, multilingual (English, Hindi, Kannada) text and voice interaction, and a clinician-in-the-loop review workflow inside a single architecture. The related-work review in Section II shows that each of these pieces has already been studied on its own, often to a high standard, but rarely combined and evaluated together within one clinically supervised system, which is the gap identified in Section III and addressed by MediAI's design in Section V. The methodology and experimental protocol in Sections VI-VII are meant to let each module eventually be judged on its own measured performance rather than on the impression created by a working end-to-end application, and Section VIII documents that the application itself, across the patient, doctor, and administrator roles, is implemented and functioning; quantitative benchmarking against that protocol is the immediate next step rather than a finished result. Put simply, MediAI's contribution at this stage isn't a claim of superior accuracy, it's the design and working implementation of an integrated, clinically supervised architecture for multimodal, multilingual AI-assisted healthcare decision support, with a clear and honest path laid out toward validating it properly.

References

[1] T. Feng, D. P. Noren, C. Kulkarni, S. Mariani, C. Zhao, E. Ghosh, D. Swearingen, J. Frassica, D. McFarlane, and B. Conroy, "Machine learning-based clinical decision support for infection risk prediction," Frontiers in Medicine, 2023.

[2] "Advancements in heart disease prediction: A machine learning approach for early detection and risk assessment," arXiv:2410.14738, 2024.

[3] W. Xue et al., "Text detection and recognition for images of medical laboratory reports with a deep learning approach," IEEE Access, vol. 8, pp. 407-416, 2020.

[4] M. W. Ma, X. S. Gao, Z. Y. Zhang, S. Y. Shang, L. Jin, P. L. Liu, F. Lv, W. Ni, Y. C. Han, and H. Zong, "Extracting laboratory test information from paper-based reports," BMC Medical Informatics and Decision Making, vol. 23, no. 1, p. 251, 2023.

[5] "High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images," Computers in Biology and Medicine, 2023.

[6] "COVID-19 detection using deep learning algorithm on chest X-ray images," PMC, 2021.

[7] R. Alkhanbouli, H. M. A. Almadhaani, F. Alhosani, and M. C. E. Simsekler, "The role of explainable artificial intelligence in disease prediction: A systematic literature review and future research directions," BMC Medical Informatics and Decision Making, vol. 25, p. 110, 2025.

[8] "Explainable AI for clinical data from EHR using SHAP and LLM-based knowledge," in Proc. Springer Conf., 2025.

[9] "Large language models-powered clinical decision support: Enhancing or replacing human expertise?" ScienceDirect, 2025.

[10] "ASR under the stethoscope: Evaluating biases in clinical speech recognition across Indian languages," arXiv:2512.10967, 2025.