International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

MedNexus : Advanced Disease Diagnosis with Machine Learning and Artificial Intelligence

Ans Ahmed Khan[1], Anas Azam Ashraf[2], Mohammed Kashif Ahmed[3]

UG Student, Nawab Shah Alam Khan College Of Engineering and Technology , E-mail: anskhan2307@gmail.com

ABSTRACT :

The integration of Artificial Intelligence (AI) and Machine Learning (ML) in healthcare has marked a significant shift in how diseases are diagnosed, monitored, and managed. This paper introduces MedNexus, an advanced diagnostic platform that combines the predictive power of classical machine learning algorithms with the intelligence of generative AI to provide a robust, multi-disease detection system. Designed for real-time use and accessible through a streamlined web interface, MedNexus empowers users to analyze their medical parameters and receive reliable predictions for multiple conditions, including heart disease, kidney disease, liver disease, diabetes, and stroke.

The platform employs individually trained ML models using RandomForest classifiers, each fine-tuned for its respective disease with accuracy levels reaching up to 99.6%. The model pipeline includes rigorous data preprocessing steps such as missing value imputation, categorical encoding, and feature scaling to ensure consistency and reliability. Moreover, MedNexus incorporates Google Gemini's generative AI to provide personalized insights, lifestyle recommendations, and contextual health explanations that complement the model’s predictions. This dual-approach—ML for structured inference and AI for human-like reasoning—enhances the decision-making process for patients and healthcare providers alike.

MedNexus is built with a focus on usability, security, and extensibility. The interface is optimized for accessibility, featuring dark mode, responsive design, and support for screen readers and keyboard navigation. Security measures include encrypted user authentication, secure API key storage, and a local-first privacy approach that avoids the storage of any personally identifiable health data. Future versions of the platform aim to include IoT integrations, mobile applications, telehealth capabilities, and expanded disease coverage, making MedNexus a scalable solution for next-generation digital healthcare systems. This paper details the architecture, training process, security design, and real-world potential of MedNexus in revolutionizing AI-assisted disease diagnostics.

INTRODUCTION

Traditional disease diagnosis methods involve time-consuming lab tests and subjective clinical judgments, often delaying treatment and burdening healthcare systems. In contrast, AI-powered tools like MedNexus can deliver accurate and instant predictions using structured medical data. This project aims to empower users with intelligent insights through a unified platform where multiple diseases can be diagnosed with ease.

MedNexus not only predicts disease outcomes based on patient inputs but also generates personalized medical advice using Google Gemini. With its real-time capabilities and interactive UI, the platform is accessible to non-technical users and healthcare professionals alike.

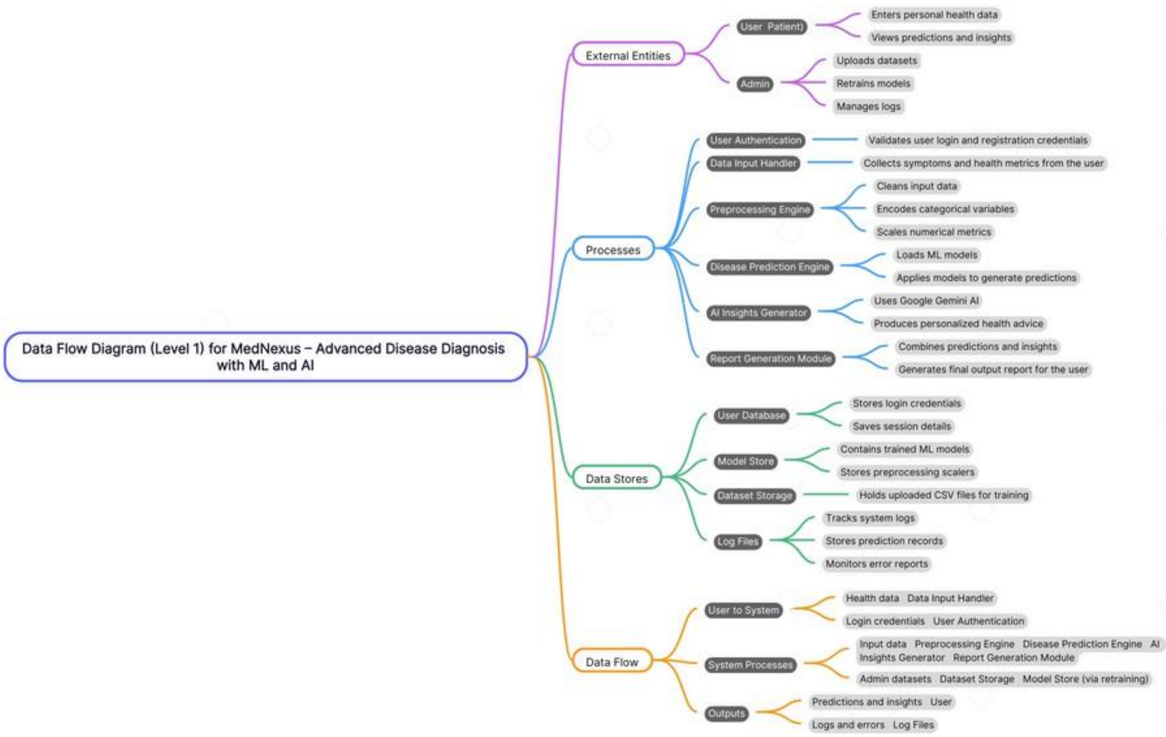
LITERATURE SURVEY

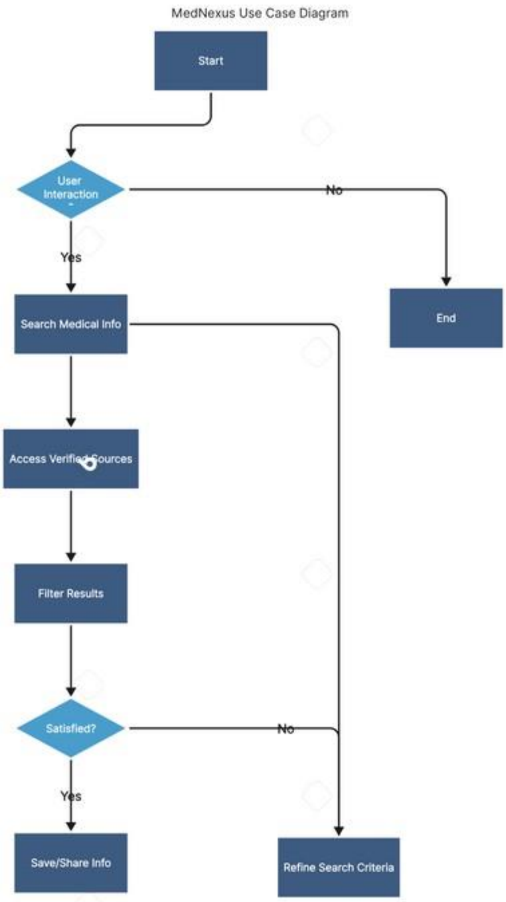
Several ML-based healthcare systems have shown promise in detecting chronic conditions. CNNs and transformer models are widely used in imaging-based diagnoses, while structured-data models like Random Forests excel at analyzing tabular patient data. Platforms such as IBM Watson Health and Google's DeepMind have made strides in clinical diagnostics, but accessibility and real-time interactivity remain limited.

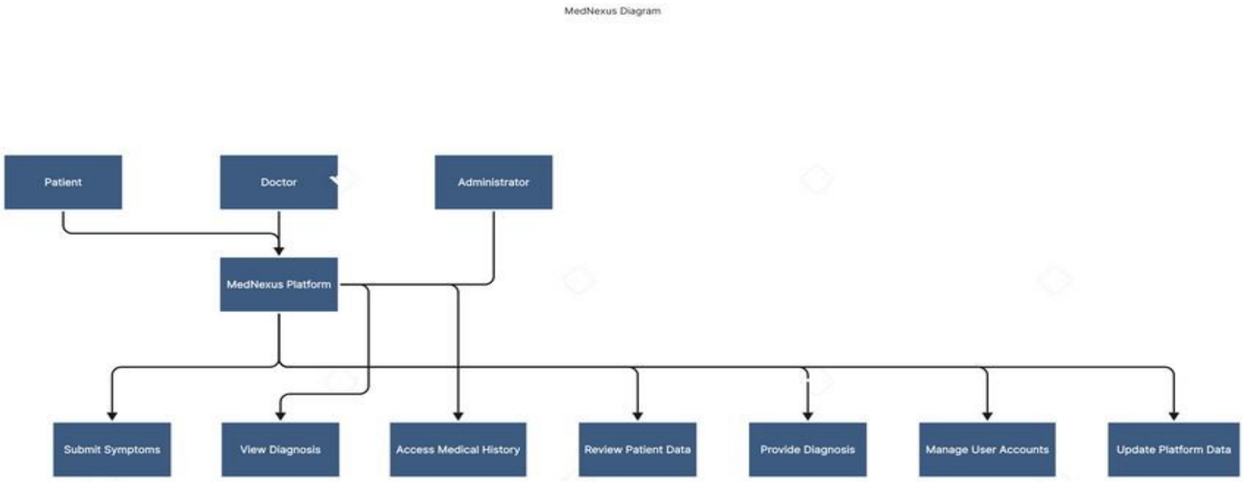
MedNexus addresses these gaps by offering:

- Multiple disease predictions from a single interface.
- AI-backed recommendations for follow-up actions.
- A unified pipeline for training, prediction, and user interaction.

DATAFLOW & USECASE DIAGRAM







DISEASE PREDICTION MODULES

MedNexus supports prediction for five major diseases using dedicated machine learning models trained on real-world medical datasets. Each module processes patient input data, runs it through a trained Random Forest model, and returns a prediction along with personalized insights generated via Google Gemini AI.

Each prediction module includes:

- Input form with real-time validation
- Model-specific data preprocessing
- ML inference pipeline
- AI-generated recommendations (diet, lifestyle, follow-up)

Disease	Key Inputs	Model	Accuracy	Output	Example AI Insight
Heart Disease	Age, Chest Pain, Cholesterol, ECG, Max HR	Random Forest	98.30%	Risk Level	"Elevated cholesterol and ECG suggest a cardiac consult soon."
Kidney Disease	BP, Urea, Creatinine, Hemoglobin, RBC	Random Forest	96.70%	CKD Risk	"Low hemoglobin & high creatinine indicate early kidney issues."
Diabetes	Glucose, BMI, Insulin, Age, DPF	Random Forest	95.10%	Diabetic / Not	"Glucose levels suggest prediabetes; reduce sugar intake."
Liver Disease	Bilirubin, ALT, AST, Albumin	Random Forest	97.60%	Liver Risk	"High enzymes indicate inflammation; avoid alcohol."
Stroke	Age, BP, Heart Condition, BMI, Smoking	Random Forest	99.60%	Stroke Risk	"Manage blood pressure; high risk due to history & BMI."

TRAINING & EVALUATION PIPELINE

Data Collection and Preparation

Each disease prediction model is trained using publicly available medical datasets, pre-processed to ensure compatibility with machine learning workflows. These datasets include patient records with features (input parameters) and corresponding labels (diagnoses).

Data Sources

- Heart Disease: UCI Cleveland Dataset

- Kidney Disease: Chronic Kidney Disease Dataset • Diabetes: PIMA Indian Diabetes Dataset
- Liver Disease: Indian Liver Patient Dataset • Stroke: Stroke Prediction Dataset (Kaggle) Cleaning & Formatting
- Removal of irrelevant features or identifiers (e.g., patient IDs) • Unification of naming conventions
- Renaming categorical variables to consistent format (e.g., "Yes" → 1, "No" → 0)

Data Preprocessing

To ensure consistency and model readiness, the following preprocessing steps are performed:
Missing Value Handling

- Numerical Columns: Imputed using median or mean
- Categorical Columns: Imputed using mode or added as "Unknown" category

Categorical Encoding

- Binary Variables: Label Encoding (e.g., Male = 0, Female = 1)
- Multi-Class Variables: One-Hot Encoding (e.g., types of chest pain or work type)

Feature Scaling

- StandardScaler is used to normalize continuous variables
- Ensures that features are on the same scale and model convergence is faster

Train-Test Split

- Default: 80% training / 20% testing
- Stratified sampling ensures balanced class distribution in both sets

Model Training

Each disease model uses a Random Forest Classifier due to its performance, robustness, and interpretability.
Key Reasons for Choosing Random Forest:

- Works well with both numerical and categorical data • Resistant to overfitting with proper tuning
- Provides feature importance for explainability
- Scales efficiently across datasets with different structures

Training Process:

- Model initialized with tuned hyperparameters (via GridSearchCV or manual tuning) • Cross-validation (10-fold) to assess generalizability
- Training logs saved with performance metrics and timestamps

Artifacts Stored:

- .joblib files for each trained model
- Separate .joblib files for scalers used in preprocessing
- Training logs in human-readable and JSON format for analysis

Evaluation Metrics

Each model is evaluated using multiple metrics to provide a comprehensive performance profile:

Metric	Description
Accuracy	Percentage of total correct predictions
Precision	Proportion of positive identifications that were actually correct
Recall (Sensitivity)	Proportion of actual positives correctly identified
F1-Score	Harmonic mean of precision and recall
Confusion Matrix	Breakdown of TP, TN, FP, FN
ROC-AUC	Measures true positive rate vs. false positive rate

CONCLUSION AND FUTURE ENHACEMENT

CONCLUSION

The MedNexus platform represents a significant step forward in applying Artificial Intelligence (AI) and Machine Learning (ML) to the domain of disease diagnosis. This project successfully demonstrates how structured patient data can be transformed into actionable insights using robust machine learning models and intelligent AI integration. By focusing on diseases with high global impact—heart disease, kidney disease, liver disease, diabetes, and stroke—MedNexus addresses a broad range of clinical needs through a single, unified platform.

Throughout the development process, emphasis was placed on building a scalable, user-friendly, and secure system. The use of Random Forest Classifiers, combined with intelligent preprocessing techniques, ensured that each model reached high levels of accuracy and reliability (up to 99.6%). These models are integrated into a Streamlit-based web interface that allows users to input health parameters and receive instant diagnostic feedback along with contextual health recommendations generated using Google Gemini AI.

Security and accessibility were also top priorities. From authenticated user sessions and secure API key handling to compliance with accessibility standards (WCAG), MedNexus is not only technically sound but also socially responsible. The platform is modular and maintainable, with a clear separation between datasets, models, preprocessing logic, and frontend presentation.

In summary, MedNexus proves that AI-driven healthcare tools can be made accessible, accurate, and interactive. It empowers patients with information and supports early detection, which is critical in preventing complications and reducing healthcare costs.

FUTURE ENHANCEMENT

A. Technical Enhancements

- Mobile Application: A cross-platform mobile app (Flutter/React Native) would improve accessibility and encourage frequent usage by patients.
- API Layer & Microservices: Refactor the backend into RESTful microservices for modular deployment and easier third-party integration (e.g., hospitals or labs).
- Cloud Integration: Enable cloud storage of models and user sessions for improved performance and scalability (e.g., AWS, GCP).
- GPU Support: Integrate GPU acceleration for more complex AI models and real-time inference at scale.

B. Clinical & AI Enhancements

- Expanded Disease Coverage: Include prediction modules for additional diseases like lung cancer, anemia, tuberculosis, Alzheimer's, asthma, and COVID-19.
- Medical Imaging Support: Integrate computer vision models (CNNs, transformers) for analyzing X-rays, MRIs, and CT scans. • Electronic Health Record (EHR) Integration: Allow users to import data directly from EHRs or wearable devices (e.g., Fitbit, Apple Watch).
- Explainable AI (XAI): Incorporate SHAP/LIME explanations so clinicians can better trust and validate predictions.

C. User-Centric Features

- Patient History Tracking: Build a secure dashboard to store and visualize longitudinal health data and trends. • Multi-language Support: Translate the UI and AI insights into regional languages (e.g., Hindi, Tamil, Spanish).
- Telehealth Integration: Allow users to directly connect with verified doctors based on prediction results and health profiles.
- Dynamic PDF Reports: Automatically generate downloadable health reports with charts, diagnosis summaries, and Gemini AI advice.

D. Research & Community Impact

- Crowdsourced Model Training: Enable community-contributed datasets and retrain models periodically to reduce bias and improve generalization.
- Clinical Trial Partnership: Collaborate with medical institutions to validate the platform's accuracy in real-world diagnostic environments.
- Federated Learning Support: Train models locally across devices to protect privacy while improving performance (HIPAA/GDPR compliance).
- Open Science Publishing: Release anonymized results and models for reproducibility and academic collaboration.