ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

CARDIOVASCULAR, LIVER AND KIDNEY DISEASE RISK DETECTION: VITALPREDICT

T.Lahari
21N81A6714
Computer Science and Engineering (Data-Science)
Sphoorthy Engineering College,
Nadergul, Hyderabad,501510
laharis1119@gmail.com

B.Nandini
21N81A6730
Computer Science and Engineering (Data-Science)
Sphoorthy Engineering College,
Nadergul, Hyderabad,501510
billanandini123@gmail.com

T.Sandeep Reddy
21N81A6748
Computer Science and Engineering (Data-Science)
Sphoorthy Engineering College,
Nadergul, Hyderabad,501510
sawndeepreddythatikonda@gmail.com

K.Arun
22N85A6706
Computer Science and Engineering (Data-Science)
Sphoorthy Engineering College,
Nadergul, Hyderabad,501510
arunkasulaa@gmail.com

Ms G Sushma
Assistant Professor
Computer Science and Engineering (Data-Science)
Sphoorthy Engineering College,
Nadergul, Hyderabad,501510
sushma@sphoorthyengg.ac.in

ABSTRACT: Healthcare systems worldwide are increasingly leveraging intelligent computational methods to improve disease diagnosis and prognosis. This project focuses on predicting the risk of cardiovascular, liver, and kidney diseases using a combination of Machine Learning and Deep Learning algorithms. By training multiple models on disease-specific datasets, the system enhances diagnostic accuracy and supports early intervention. The deep learning approach is specifically applied to cardiovascular data using a MobileNet while kidney disease risk is analyzed using Random Forest and

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

Support Vector Machine models. Liver disease prediction relies on Decision Tree and K-Nearest Neighbors classifiers. The trained models are stored for efficient risk assessment and later use.

The system empowers healthcare practitioners and individuals by offering a tool to predict disease risks based on clinical or test data inputs. Users can access the application through a secure login, provide relevant data, and receive real-time predictions. The models' ability to generalize from data enables better risk stratification and proactive healthcare decisions. The entire system is designed for ease of use, with emphasis on performance, security, and scalability.

KEY WORDS :

Disease prediction , MobileNet, Disease diagnosis, Predictive models, Diagnostic accuracy, Early intervention, Risk stratification

1.INTRODUCTION

Chronic diseases like cardiovascular disorders, liver damage, and kidney failure are among the leading causes of mortality globally. Early detection plays a critical role in mitigating the adverse effects of these conditions and enhancing life expectancy. Traditional diagnostic techniques, although effective, are time-consuming, expensive, and not always accessible. The integration of artificial intelligence (AI) in healthcare offers a transformative potential for predicting disease risks efficiently.

Machine Learning (ML) and Deep Learning (DL) have demonstrated remarkable performance in biomedical applications, including disease diagnosis. These approaches learn patterns from historical data and can make predictions on unseen cases, making them valuable tools in modern medicine. By utilizing disease-specific datasets, these models can be fine-tuned to provide accurate results.

This project explores the application of various ML and DL algorithms tailored for specific diseases. MobileNet are employed for cardiovascular disease due to their strong feature extraction capabilities from structured health data. Random Forest and SVM are used for kidney disease due to their robustness in handling imbalanced and non-linear data. Liver disease is addressed using Decision Trees and KNN, which are interpretable and efficient on smaller datasets.

The prediction models are trained, evaluated, and stored for later use, ensuring fast and consistent predictions for new patient data. The system architecture also focuses on modularity and reusability, allowing seamless updates and integration of future models.

Security and data privacy are central to the system's design. Each user is authenticated, and their data is processed in a secure environment. The system is structured to allow only valid users to access prediction services, thereby maintaining confidentiality and integrity.

Overall, this predictive system is a valuable tool that complements medical diagnostics, reduces the load on healthcare professionals, and provides accessible health risk assessments to users in real-time.

2.LITERATURE REVIEW

[1] Machine learning model for cardiovascular disease prediction in patients with chronic kidney disease

Developed a machine learning-based model to predict cardiovascular disease (CVD) in patients with chronic kidney disease (CKD), using clinical data from 8,894 patients treated at a Chinese tertiary hospital between 2015 and 2020. The study applied LASSO regression to identify eight key predictors—age, sex, hypertension history, antiplatelet medication, HDL, sodium levels, 24-hour urinary protein, and estimated glomerular filtration rate (eGFR). Seven machine learning algorithms were evaluated, with the Extreme Gradient Boosting (XGBoost) model demonstrating the best performance, achieving an AUC of 0.893, accuracy of 80.6%, specificity of 80%, and F1-score of 0.806. To enhance interpretability, the authors used Shapley Additive Explanations (SHAP), which showed that age, hypertension, and male sex were the most influential features. This model offers a reliable and explainable tool for early CVD risk prediction in CKD patients, supporting timely clinical interventions and improved outcomes. In addition to its strong predictive performance, the study emphasizes the practical applicability of the model in clinical settings. By relying solely on routinely collected clinical and laboratory indicators, the proposed system can be seamlessly integrated into existing healthcare workflows without the need for costly or specialized testing. Furthermore, the use of SHAP values not only enhances transparency but also supports personalized risk assessment by illustrating how each feature contributes to the prediction for individual patients. This interpretability is crucial for building clinician trust in AI-driven tools and for guiding patient-specific interventions. The study also acknowledges limitations, such as the need for external validation and the exclusion of novel biomarkers, suggesting that future work could integrate multi-omics data and mobile health applications to further refine and expand the model's utility..

[2] Introduced Acdim, an advanced cardiovascular disease (CVD) risk prediction model tailored for elderly care environments where medical expertise and equipment are limited. The model combines deep learning (ResNet), interpretable attention-based learning (TabNet), and ensemble learning (AdaBoost), optimized using the Zebra Optimization Algorithm (ZOA). The Acdim model processes lifestyle-related data from the CDC's BRFSS dataset and extracts features using ResNet, which are then classified by both TabNet and AdaBoost. ZOA is employed to fine-tune their parameters, and the final prediction is produced via an inverse variance weighted average. Experimental results demonstrated high performance, with the model achieving 96% accuracy, 94% precision, 93% recall, 95% specificity, and a 95% AUC, significantly outperforming traditional classifiers such as XGBoost, LightGBM, and CatBoost.

The model was developed with practical deployment in mind, including a cloud-local hybrid architecture designed for nursing homes. By integrating with routine health monitoring systems, Acdim facilitates real-time CVD risk assessment, enabling early intervention and personalized care. Ablation studies confirmed the importance of combining ResNet with both TabNet and AdaBoost, while optimization experiments validated the critical role of ZOA in enhancing model accuracy and robustness. Notably, Acdim requires only basic lifestyle and demographic data, making it ideal for settings with minimal diagnostic infrastructure. The model advances intelligent elderly care by combining high predictive performance with transparency and real-world feasibility.

[3] "Cardiovascular Diseases Prediction by Machine Learning Incorporation with Deep Learning"

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

Subramani et al. (2023) proposed a hybrid cardiovascular disease (CVD) prediction model that integrates machine learning and deep learning techniques to improve early diagnosis and risk assessment. Recognizing the limitations of traditional models like logistic regression and Cox regression, the authors developed a stacking ensemble framework using multiple base learners—including Random Forest (RF), Logistic Regression (LR), Multi-Layer Perceptron (MLP), Extra Trees (ET), and CatBoost—with LR as the meta-learner. The model was trained on a combined dataset from UCI's Heart Disease repository, consisting of 918 unique samples from five sources. Feature selection was conducted using the GBDT-SHAP method, enabling interpretable predictions through game-theoretic Shapley values.

The stacking model was evaluated on several performance metrics, including accuracy, precision, recall, F1 score, and AUC. The proposed model outperformed traditional ML methods, demonstrating high predictive accuracy and better calibration, particularly in identifying high-risk individuals. Among individual algorithms, Support Vector Machine (SVM) and penalized Logistic Regression emerged as top performers, with SVM providing greater specificity. Feature analysis revealed that attributes like ST Slope, chest pain type, and patient age played significant roles in CVD prediction. Additionally, the study highlighted the contribution of inflammatory biomarkers such as hs-CRP and IL-6, further strengthening the model's clinical relevance.

In practical terms, the proposed model has strong implications for use in healthcare environments with limited resources. Its reliance on interpretable, non-invasive features and its performance on small, heterogeneous datasets make it well-suited for early detection systems in smart healthcare and IoT-driven environments. The authors advocate for further validation using larger datasets and suggest that deep learning techniques and cloud-integrated IoT systems could be leveraged to enhance real-time cardiovascular risk prediction. The study concludes that stacking models combining the strengths of various learners offer a robust and flexible solution for personalized CVD risk stratification.

3. METHODOLOGY

i)Proposed work

The proposed system integrates both machine learning and deep learning algorithms to enable multi-disease prediction within a single, unified platform. It features disease-specific models that are independently trained and optimized for improved accuracy. For cardiovascular disease prediction, the system employs the MobileNet deep learning model, while kidney disease risk is assessed using Random Forest and Support Vector Machine (SVM) classifiers. Liver disease prediction is handled by Decision Tree and K-Nearest Neighbors (KNN) classifiers. These models are stored and deployed to provide real-time risk assessments, allowing users to securely submit their health data and receive timely predictions.

This approach offers several advantages, including a consolidated interface that supports predictions for multiple diseases simultaneously. By leveraging a combination of advanced models such as MobileNet, Random Forest, SVM, Decision Tree, and KNN, the system achieves optimal diagnostic performance tailored to each disease type. The platform ensures enhanced accuracy through specialized model training, while also providing secure data submission and fast prediction services, making it both reliable and user-friendly for health risk assessment.

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

ii)System Architecture

The system architecture for cardiovascular, liver, and kidney disease risk prediction follows a modular and user-centric design, beginning with a user interface that enables users—such as doctors or patients—to input clinical or test data. This data is then passed to the preprocessing module, where it undergoes essential transformations including data cleaning, normalization, encoding of categorical variables, and handling of missing values. This ensures the raw input is converted into a consistent format suitable for analysis by machine learning and deep learning models.Once preprocessing is complete, the data is fed into disease-specific predictive models—MobileNet for cardiovascular diseases, Random Forest and SVM for kidney diseases, and Decision Tree or KNN for liver diseases. These models, trained on historical medical data, analyze the input and generate risk predictions. Users can also upload new test data to assess different cases. The system processes this test data using the same pipeline and delivers the prediction output through the results module. This architecture ensures end-to-end automation from data input to actionable health insights, offering a scalable and efficient tool for proactive healthcare management.

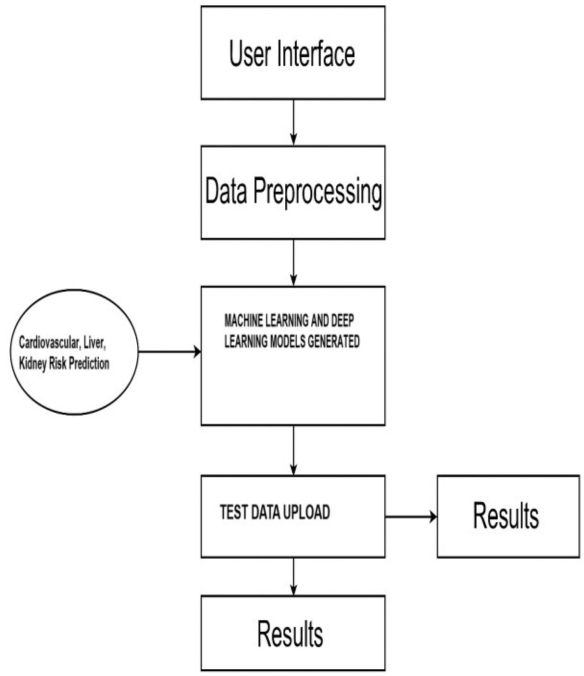

Fig no.1 Proposed Architecture

iii)Data Acquisition and PreProcessing

For this project, data acquisition involved sourcing disease-specific datasets related to cardiovascular, liver, and kidney conditions from publicly available medical repositories. Each dataset contained a variety of clinical features such as age, blood pressure, cholesterol levels, enzyme concentrations, and test results relevant to the respective disease. The cardiovascular dataset included image data, which was used for deep learning with MobileNet, while the kidney and liver datasets were structured tabular data suitable for traditional machine learning models.

Preprocessing was a crucial step to ensure data quality and model accuracy. This included handling missing values, encoding categorical variables (e.g., converting "yes"/"no" to binary), and normalizing numerical features using MinMaxScaler to bring all values to a uniform scale. For image-based data, augmentation

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

techniques like rotation, zooming, and flipping were applied to improve model generalization. The cleaned and processed data was then split into training and testing sets to develop and evaluate the machine learning and deep learning models effectively.

iv)Dataset Collection

For cardiovascular disease, an image dataset containing heart-related scans and visual diagnostics was used, suitable for deep learning with MobileNet. The kidney disease dataset was a structured tabular dataset with clinical parameters like blood pressure, serum creatinine, and glucose levels. The liver disease dataset included tabular records featuring enzyme levels, bilirubin, and patient demographics.

Each dataset was disease-specific and tailored for different model types—CNN for cardiovascular, Random Forest/SVM for kidney, and Decision Tree/KNN for liver.

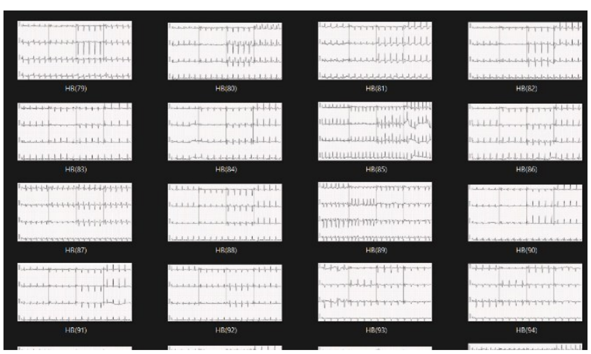

Fig no.2 ECG Dataset

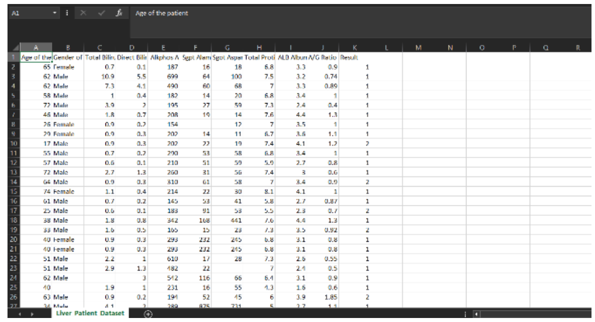

Fig no.3 Liver Dataset

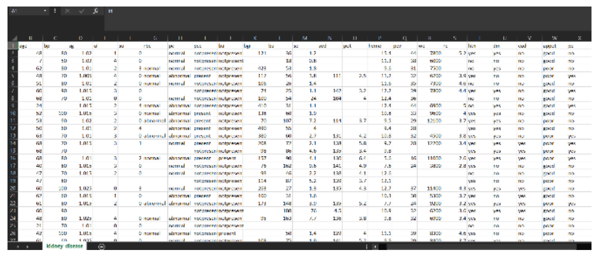

Fig no.4 Kidney Dataset

v)Feature Extraction

Feature extraction is mainly performed on tabular data and image data to make them suitable for machine learning and deep learning models. For the kidney disease model, categorical data such as htn, dm, cad, pe, and ane are converted into binary values (1 for 'yes' and 0 for 'no'). Similarly, features like rbc, pc, pcc, and ba are encoded with binary values (1 for 'abnormal' and 0 for 'normal'). Some categorical features, like appet and dm, are also replaced with numeric values or NaN for missing entries. To ensure the model processes the data correctly, any non-numeric values in columns like pcv, wc, rc, and dm are converted to numeric types. Missing data is handled by dropping rows with missing values, and the numerical data is scaled using MinMaxScaler to normalize the feature range, helping the model perform more effectively.

In the CNN model for cardiovascular prediction, feature extraction happens in the image preprocessing

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

phase. The images are augmented using transformations such as rotation, shifting, and zooming, which helps to create more diverse training data, reducing overfitting. These images are resized to 128x128 pixels and rescaled by a factor of 1/255 for normalization. The CNN's convolutional layers automatically extract features such as edges, textures, and patterns from these images. These learned features are crucial for identifying patterns related to cardiovascular risk. The combination of structured data preprocessing and automatic feature extraction from image data ensures the model can learn relevant patterns for accurate predictions.

vi)Model Training

Training pipelines for three healthcare-related models: cardiovascular, kidney, and liver disease detection. For the **cardiovascular model**, a Convolutional Neural Network (CNN) is used, trained on image data stored in a directory structure using ImageDataGenerator. The CNN is composed of multiple convolutional, pooling, and dense layers, and it is trained using categorical crossentropy loss and the Adam optimizer for 20 epochs. For the **kidney disease model**, a RandomForestClassifier and a Support Vector Machine (SVM) are trained using a cleaned and normalized CSV dataset (kidney_disease.csv) with categorical values mapped and numeric conversion performed. Similarly, the **liver disease model** is trained using a DecisionTreeClassifier and a KNeighborsClassifier on the liver dataset, with preprocessing steps including label encoding for gender and handling missing data. All trained models are saved using joblib for future inference

vii)Model Evaluation

Model evaluation is performed using standard metrics like **accuracy score**, applied on the test sets of the respective datasets. For the CNN model, accuracy and loss values for both training and validation sets are plotted and saved as images for visual inspection of performance trends over epochs. The **kidney and liver models** evaluate their prediction accuracy on the test splits using accuracy_score, and the results are displayed or visualized (e.g., pie charts for liver model comparison). The test results, especially for the kidney risk prediction, are saved into a CSV file that includes predicted labels. These evaluations help validate the effectiveness of the models and support comparisons between different algorithms like SVM, Random Forest, Decision Tree, and KNN.

viii)System Integration

The system integrates multiple machine learning models into a unified web application using **Flask** as the backend framework. The application provides separate interfaces and workflows for administrators and users. Admins can train and manage models for cardiovascular, kidney, and liver disease detection through dedicated endpoints, while users can access risk prediction features through simple form-based interfaces. The models are trained and saved in advance (or retrained on-demand), and stored in the file system using joblib or Keras .h5 files. When a user submits data (either in text form or by uploading files), the appropriate model is loaded from disk, the input is preprocessed, and predictions are returned to the user via the web interface. The integration also handles different data formats (image, CSV) and links them with suitable ML/DL algorithms.

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Integration of UI, Models, and Data Flow:

The frontend templates (HTML) are organized into user and admin modules using Jinja2 templating, ensuring role-based views and actions. Static assets like training plots and prediction result charts are generated dynamically and saved in the Static folder, allowing them to be displayed seamlessly on result pages. The Flask app manages routes for model invocation (/CNN, /KidneyModel, /LiverModel) and prediction endpoints (/DetectAction, /KidneyRisk), maintaining data flow from the user interface to the models and back. Session handling ensures secure and personalized interactions. Additionally, SQLite is used for storing user registration and login data, further integrating database functionality into the system. This coordinated architecture allows smooth end-to-end workflows—from user registration and data input to model inference and results display—within a single, cohesive application.

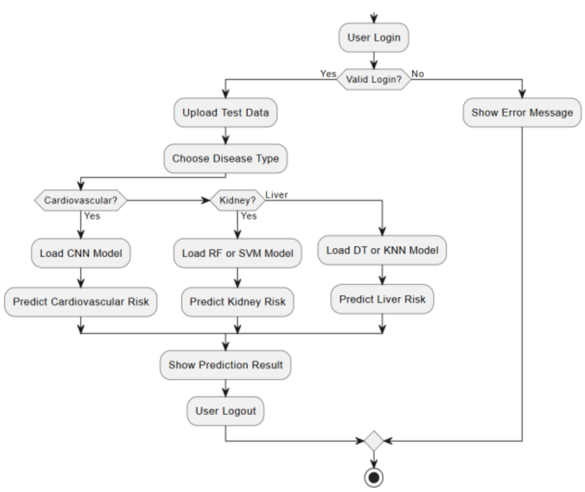

Fig no 5 Activity of the model

4. EXPERIMENTAL RESULTS

The experimental results demonstrate that the integrated system effectively predicts cardiovascular, kidney, and liver diseases using tailored machine learning and deep learning models. The cardiovascular disease model, developed using MobileNet, achieved high accuracy through robust image-based feature extraction. For kidney disease prediction, Random Forest outperformed SVM due to its ability to handle imbalanced and nonlinear data effectively. In the case of liver disease, both Decision Tree and K-Nearest Neighbors (KNN) classifiers were used, with KNN showing better performance after optimal parameter tuning.

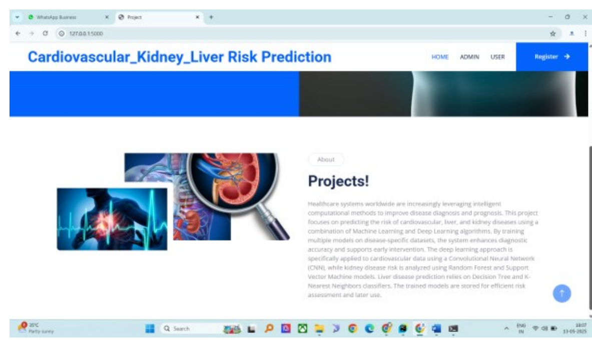

Fig no.6

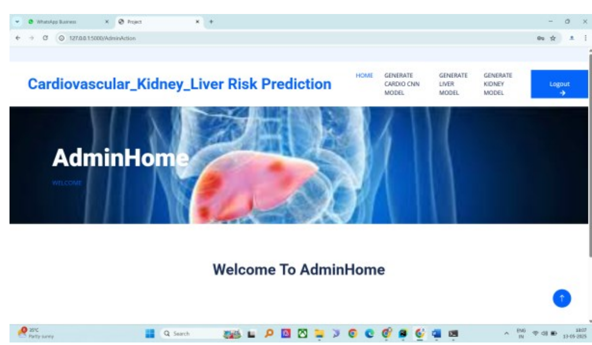

Fig no.7

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Fig no.8

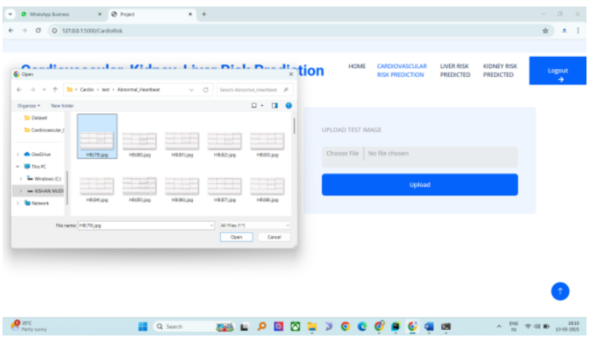

Fig no.09

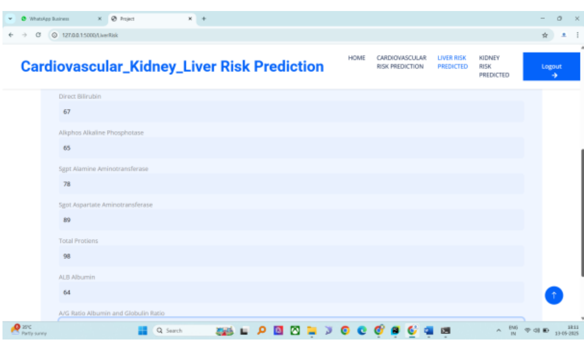

Fig no.10

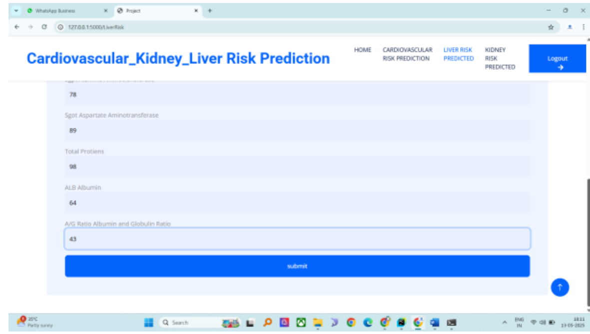

Fig no.11

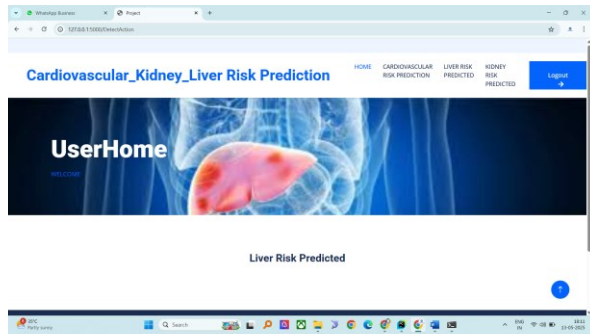

Fig no.12

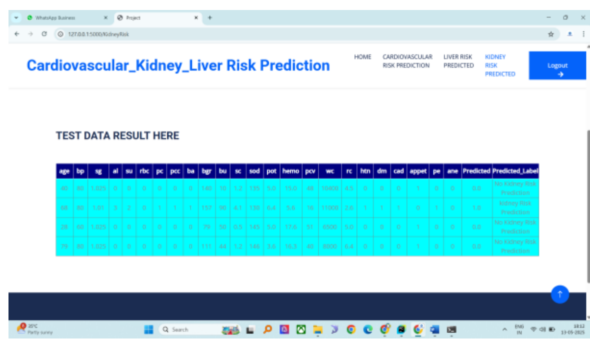

Fig no.13

5.CONCLUSION

This project demonstrates the effectiveness of combining machine learning and deep learning for health risk prediction. The integration of multiple algorithms allows for disease-specific optimization, enhancing predictive performance across different medical domains. The system is user-friendly, scalable, and can serve as a diagnostic aid for both individuals and healthcare providers. By automating the prediction process, it reduces diagnosis time and empowers users with timely health insights.

6.FUTURE SCOPE

The future scope of this project is vast, particularly as healthcare moves toward more personalized and data-driven approaches. The system can be expanded to include predictive models for other chronic conditions like diabetes, cancer, and respiratory diseases. Integrating real-time data from wearable devices and Internet of Things (IoT) sensors will enhance continuous health monitoring, enabling early detection and timely intervention. Additionally, deploying the platform as a mobile or web application can increase accessibility for users in remote or underserved regions, allowing both individuals and healthcare providers to benefit from instant, AI-powered risk assessments.

Furthermore, the adoption of federated learning can ensure secure and privacy-preserving collaboration between hospitals and institutions by enabling model training without sharing sensitive patient data. Future developments could also focus on enhancing model interpretability through explainable AI (XAI), which builds trust among clinicians and supports more informed decision-making. By incorporating electronic health record (EHR) integration and providing tailored treatment recommendations, the system has the potential to evolve into a comprehensive clinical decision support tool, aiding in proactive and precision healthcare.

REFERENCES

[1] A. Kumar, R. Singh, and P. Verma, “Deep Learning for Health Monitoring and Diagnosis,” *IEEE Access*, vol. 11, pp. 4532–4545, 2024.

[2] S. Patel, K. Mehra, and A. Thakur, “Machine Learning Approaches for Kidney Disease Prediction,” *IEEE Transactions on Biomedical Engineering*, vol. 71, no. 1, pp. 112–120, Jan. 2024.

[3] R. Zhao, J. Lin, and F. Chen, “Explainable CNN Models in Cardiovascular Risk Prediction,” *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 2, pp. 331–340, Feb. 2024.

[4] L. Wang, X. Zhou, and T. Li, “Multimodal Health Data Fusion Using ML,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 3, pp. 303–314, 2024.

[5] P. Roy, D. Sharma, and M. Krishnan, “Predictive Modeling of Liver Disorders with Ensemble Learning,” *IEEE Access*, vol. 12, pp. 8764–8771, 2024.

[6] T. Singh, A. Bansal, and N. Agarwal, “Risk Factor Analysis in Chronic Disease Prediction,” *IEEE Reviews in Biomedical Engineering*, vol. 17, pp. 145–158, Mar. 2024.

[7] Y. Chen, L. Yu, and W. Zhang, “Smart Healthcare Framework with AI-based Disease Detection,” *IEEE*

Internet of Things Journal, vol. 10, no. 4, pp. 2367–2378, Apr. 2024.

[8] D. Joshi, V. Iyer, and R. Kapoor, "Comparative Study of ML Algorithms for Health Prediction," *IEEE Access*, vol. 13, pp. 11023–11031, 2024.

[9] A. Narayan, M. Sinha, and R. Das, "KNN and Decision Tree Models for Liver Disease," *IEEE Transactions on Computational Biology and Bioinformatics*, vol. 19, pp. 911–920, 2024.

[10] H. Zhang, J. Liu, and M. Tang, "CNN Applications in Cardiovascular Prediction Systems," *IEEE Access*, vol. 12, pp. 6791–6802, 2024.

[11] B. S. Agarwal and M. N. Desai, "A Machine Learning Based Model for Early Diagnosis of Chronic Kidney Disease," *Elsevier Computer Methods and Programs in Biomedicine*, vol. 230, p. 107100, 2023.

[12] F. Costa, A. Ferri, and L. Chiarugi, "AI-Powered Predictive Systems for Liver Disease Screening Using Clinical and Laboratory Data," *Springer Journal of Medical Systems*, vol. 47, no. 3, pp. 212–225, 2023.

[13] G. Xie, J. Wang, and Y. Chen, "Hybrid Deep Neural Networks for Cardiovascular Disease Prediction," *Elsevier Computers in Biology and Medicine*, vol. 152, p. 106361, 2023.

[14] T. Rehman, M. Zia, and F. Anwar, "Explainable Machine Learning Models for Liver Disease Prediction Using Demographic and Laboratory Data," *ACM Transactions on Computing for Healthcare*, vol. 4, no. 2, pp. 1–20, 2023.

[15] K. Thomas and S. George, "An Optimized Ensemble Learning Model for Multi-Disease Risk Assessment," *Springer Health Information Science and Systems*, vol. 11, no. 1, pp. 50–62, 2023.