International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

AI DISEASE PREDICTION SYSTEM

Gouri Shankar Behera
Department of Computer Science and Engineering (Artificial Intelligence)
GIFT Autonomous, Bhubaneswar, Odisha, India

Jyotiprakash Behera
Department of Computer Science and Engineering (Artificial Intelligence)
GIFT Autonomous, Bhubaneswar, Odisha, India

ABSTRACT

The AI Disease Prediction System is an intelligent healthcare platform developed using Artificial Intelligence, Machine Learning, and Deep Learning technologies to improve disease diagnosis accuracy and healthcare accessibility. Traditional healthcare systems often suffer from delayed diagnosis, shortage of specialists, high consultation costs, and human errors. The proposed system addresses these problems by providing automated disease prediction based on symptoms and medical image analysis.

The system integrates ensemble Machine Learning algorithms such as Random Forest, Support Vector Machine (SVM), and Gradient Boosting for symptom-based disease prediction. Additionally, a Convolutional Neural Network (CNN) based on ResNet-50 architecture is used for medical image classification including chest X-rays, skin lesion images, and retinal images. The platform also includes secure authentication, prediction history management, healthcare API integration, and an admin dashboard for monitoring and analytics.

The proposed system achieved 89.1% accuracy in symptom-based disease prediction and 91.2% accuracy in image classification. The platform improves healthcare accessibility, reduces diagnosis time, minimizes manual effort, and enhances prediction reliability. The system demonstrates how Artificial Intelligence can transform modern healthcare systems through intelligent automation and predictive analytics.

Keywords: Artificial Intelligence, Disease Prediction, Machine Learning, Deep Learning, CNN, Healthcare, Medical Image Analysis, Flask, Python.

1. INTRODUCTION

Healthcare is one of the most important sectors in modern society. Early disease detection and accurate diagnosis are essential for effective treatment and patient care. However, traditional healthcare systems face several challenges including delayed diagnosis, shortage of healthcare professionals, high medical expenses, and inconsistent diagnostic accuracy.

In many rural and remote areas, patients do not have access to specialized doctors and advanced diagnostic facilities. As a result, diseases are often detected at later stages, increasing treatment complexity and mortality rates. Human errors and manual diagnosis processes also affect healthcare quality.

Artificial Intelligence (AI) has emerged as a powerful technology capable of solving many healthcare-related problems. AI-based systems can analyze large amounts of medical data, identify patterns, and generate accurate predictions within seconds. Machine Learning algorithms are widely used in healthcare applications for disease prediction, patient monitoring, medical image analysis, and treatment recommendations.

he proposed AI Disease Prediction System combines Machine Learning and Deep Learning technologies into a unified healthcare platform. The system predicts diseases based on user symptoms and analyzes medical images using Convolutional Neural Networks (CNNs). The platform provides secure login, prediction history, disease information, and healthcare analytics.

The main goal of the system is to improve healthcare accessibility, reduce diagnosis time, and provide accurate disease prediction support for patients and healthcare professionals.

2. OBJECTIVES OF THE PROJECT

The major objectives of the AI Disease Prediction System are:

1. To develop an intelligent disease prediction system using Artificial Intelligence.
2. To predict diseases based on user symptoms using Machine Learning algorithms.
3. To analyze medical images using Deep Learning techniques.
4. To improve diagnosis accuracy and reduce medical errors.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

5. To provide secure user authentication and data privacy.
6. To maintain prediction history and healthcare records.
7. To integrate healthcare APIs for disease information.
8. To reduce healthcare costs and improve accessibility.
9. To support responsive web-based healthcare services.
10. To provide an admin dashboard for monitoring and analytics.

The project mainly focuses on integrating Machine Learning and Computer Vision technologies into healthcare systems for accurate and efficient disease diagnosis.

3. LITERATURE SURVEY

Artificial Intelligence and Machine Learning technologies are increasingly being used in modern healthcare systems. Researchers have developed various disease prediction models using supervised learning algorithms and deep learning techniques.

Machine Learning models such as Decision Trees, Random Forest, Naive Bayes, and Support Vector Machines are commonly used for predicting diseases from symptoms and patient medical records. These algorithms analyze symptom patterns and generate disease predictions with high accuracy.

Deep Learning models such as Convolutional Neural Networks (CNNs) are widely used in medical image analysis. CNN-based systems can detect diseases from X-rays, MRI scans, retinal images, and skin lesion photographs. Pretrained models such as ResNet, VGG, and Inception are commonly used for transfer learning in healthcare applications.

Several online healthcare platforms provide symptom-based diagnosis support. However, many existing systems lack integrated medical image analysis, explainable AI, and secure healthcare management features.

The proposed AI Disease Prediction System combines symptom prediction and image analysis into a single platform, improving healthcare efficiency, accessibility, and reliability.

4. EXISTING SYSTEM

Traditional healthcare systems depend mainly on manual diagnosis and clinical expertise. Patients are required to visit hospitals physically, undergo multiple tests, and wait for specialist consultations. This process consumes significant time and financial resources.

Existing online healthcare platforms provide symptom-based suggestions but often lack advanced AI capabilities. Many systems do not support image analysis or secure patient record management.

Some major limitations of existing systems are:

Delayed diagnosis and treatment

Limited healthcare accessibility in rural areas

High consultation and testing costs Human errors in diagnosis

Lack of automated medical image analysis Poor healthcare data management

Limited prediction accuracy

Lack of integrated healthcare analytics

Most existing systems focus either on symptom prediction or image analysis separately. Therefore, there is a need for an integrated AI-based healthcare platform capable of providing secure and accurate disease diagnosis.

5. PROPOSED SYSTEM

The proposed AI Disease Prediction System integrates Machine Learning, Deep Learning, and healthcare APIs into a unified platform.

The system includes the following major modules:

User Module Admin

Module

Symptom Prediction Module Medical Image Analysis Module Database Management Module Authentication and Security Module

During symptom prediction, users select symptoms from a predefined symptom list. The selected symptoms are processed using Machine Learning models including Random Forest, SVM, and Gradient Boosting.

For image analysis, users upload medical images such as X-rays, skin lesion images, or retinal scans. The system processes these images using a CNN model based on ResNet-50 architecture.

The system also includes secure login, JWT authentication, encrypted password storage, prediction history management, and healthcare API integration.

The proposed system improves healthcare accessibility, diagnosis accuracy, and operational efficiency while reducing manual effort and healthcare costs

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

6. SYSTEM REQUIREMENTS

6.1 Hardware Requirements

- Intel Core i3 Processor or Higher
- 4 GB RAM
- 500 GB Hard Disk
- Internet Connection
- Webcam (Optional for image upload)

6.2 Software Requirements

- Operating System: Windows/Linux
- Programming Language: Python
- Frontend: HTML, CSS, JavaScript, Bootstrap
- Backend: Flask
- Database: SQLite
- Libraries: TensorFlow, Keras, OpenCV, NumPy, Pandas, Scikit-learn

7. SYSTEM ARCHITECTURE

The AI Disease Prediction System follows a layered architecture model consisting of:

1. Presentation Layer
2. Application Layer
3. AI/ML Layer
4. Data Layer

The frontend interface allows users to register, log in, select symptoms, upload images, and view prediction results.

The backend layer processes user requests and communicates with Machine Learning models and the database.

The AI/ML layer contains:

- Random Forest
- SVM
- Gradient Boosting
- CNN ResNet-50

The database layer stores:

- User information
- Disease records
- Symptoms
- Prediction history
- Image analysis result

The architecture ensures scalability, security, and efficient healthcare data management

8. DATA FLOW DIAGRAM

The Data Flow Diagram (DFD) represents the movement of data within the AI Disease Prediction System. It explains how information flows between users, Machine Learning models, databases, and system modules. The DFD helps in understanding the working process of the system in a simple and structured manner.

The AI Disease Prediction System mainly consists of the following components:

1. User
2. Input Validation Module
3. Preprocessing Module
4. Machine Learning Prediction Module
5. CNN Image Analysis Module
6. Database Management System
7. API Integration Module
8. Admin Module

The data flow process begins when the user enters symptoms or uploads medical images into the system. The user interacts with the web interface developed using HTML, CSS, Bootstrap, and JavaScript. The input may include symptoms such as fever, headache, cough, fatigue, or medical images like X-rays and skin lesion photographs.

After receiving the input, the system sends the data to the Input Validation Module. This module checks whether the entered symptoms and uploaded images are valid or not. It removes incorrect or incomplete data and ensures proper formatting before further processing.

Once validation is completed, the data moves to the Preprocessing Module. In this stage, symptom data is converted into feature vectors suitable for Machine Learning algorithms. For medical images, preprocessing operations such as resizing, normalization, grayscale conversion, and noise removal are performed. These preprocessing techniques improve prediction accuracy and model performance.

The processed symptom data is then transferred to the Machine Learning Prediction Module. This module uses ensemble Machine Learning algorithms such as Random Forest, Support Vector Machine (SVM), and Gradient Boosting to predict diseases. Each model generates prediction probabilities, and the ensemble method combines these results using weighted voting techniques to produce the final disease prediction.

If the user uploads a medical image, the image is processed by the CNN Image Analysis Module. This module uses a Convolutional Neural Network based on ResNet-50 architecture for disease classification. The CNN model extracts important image features automatically and predicts diseases such as pneumonia, tuberculosis, melanoma, diabetic

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

retinopathy, and glaucoma.

After prediction, the system stores user information, prediction results, image analysis reports, and timestamps in the Database Management System. SQLite database is used for secure and efficient data storage. The database maintains prediction history, user records, diseases, symptoms, and image analysis logs.

The API Integration Module connects the system with external healthcare APIs such as OpenFDA and MedlinePlus. This module retrieves additional medical information including disease descriptions, precautions, medications, and healthcare recommendations. The fetched information is displayed along with prediction results.

Finally, the Prediction Results Module displays the predicted disease name, confidence score, precautions, and recommendations to the user through the web interface. Users can also view their previous prediction history securely.

The Admin Module continuously monitors system activities. Administrators can manage users, monitor predictions, update disease databases, and analyze healthcare statistics using analytics dashboards.

The DFD ensures proper understanding of how data flows between different components of the AI Disease Prediction System. It improves system transparency, simplifies development.

9. DATABASE DESIGN

The database stores all system-related information securely.

Major tables include:

- Users Table
- Diseases Table
- Symptoms Table
- Predictions Table
- Disease-Symptom Mapping Table
- Image Analysis Table

The database maintains healthcare records, prediction logs, and user activity history. SQLite is used due to its lightweight architecture and fast query processing.

Security mechanisms such as encrypted passwords, authentication controls, and secure database connectivity are implemented to protect healthcare information.

10. MODULE DESCRIPTION

10.1 User Module

This module allows users to:

- Register accounts
- Log in securely
- Select symptoms
- Upload medical images
- View prediction results
- Access prediction history

10.2 Symptom Prediction Module

This module uses ensemble Machine Learning algorithms to predict diseases based on symptoms.

The models used are:

- Random Forest
- SVM
- Gradient Boosting

The final prediction is generated using weighted voting techniques.

10.3 Image Analysis Module

This module uses CNN-based Deep Learning models for image classification.

Supported image categories include:

- Chest X-rays
- Skin lesion images
- Retinal scans

10.4 Admin Module

The admin module manages:

- User accounts
- Disease databases
- Prediction monitoring
- Analytics dashboards
- Model management

implemented to protect healthcare information.

11. IMPLEMENTATION

The system is implemented using Python and Flask framework. HTML, CSS, Bootstrap, and JavaScript are used for frontend development.

Machine Learning models are developed using Scikit-learn while Deep Learning models are implemented using TensorFlow and Keras.

The symptom prediction process includes:

1. Symptom selection
2. Feature vector generation
3. Model prediction
4. Probability ranking
5. Result display

The image analysis process includes:

1. Image upload
2. Image preprocessing

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

CNN Accuracy

3. CNN inference
4. Disease classification
5. Prediction output

The backend server communicates with the database and healthcare APIs to retrieve disease-related information.

12. ALGORITHMS USED

12.1 Random Forest Algorithm

Random Forest is an ensemble learning algorithm that combines multiple decision trees for accurate classification.

Random Forest Accuracy
87.2%

12.2 Support Vector Machine (SVM)

SVM is used for classification by identifying optimal hyperplanes between disease categories.

SVM Accuracy

84.6%

12.3 Gradient Boosting

Gradient Boosting improves prediction performance using sequential learning methods.

Gradient Boosting Accuracy

86.4%

12.4 CNN ResNet-50

The CNN model processes medical images for disease classification.

91.2%

13. RESULTS AND DISCUSSION

The AI Disease Prediction System successfully achieved high accuracy in both symptom-based and image-based disease prediction.

Performance Results

Parameter	Result
Symptom Prediction Accuracy	89.1%
CNN Image Accuracy	91.
2% Average Response Time	1.8
sec Image Processing Time	6.4
sec Concurrent Users Supported	49
0 Database Query Speed	21

0 ms

The ensemble Machine Learning model improved overall prediction accuracy compared to individual models.

The CNN model performed efficiently for chest X-ray, skin lesion, and retinal image classification tasks.

The system demonstrated fast response time, secure healthcare management, and reliable disease prediction performance

14. ADVANTAGES OF THE SYSTEM

1. Provides accurate disease prediction
2. Reduces diagnosis time
3. Supports medical image analysis
4. Improves healthcare accessibility
5. Reduces healthcare costs
6. Provides secure healthcare management
7. Maintains prediction history
8. Supports real-time prediction
9. User-friendly interface

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

10. Scalable and efficient architecture

15. FUTURE ENHANCEMENTS

Future improvements include:

1. Explainable AI integration using SHAP and LIME
2. Mobile application support
3. Cloud deployment
4. Wearable device integration
5. Telemedicine consultation support
6. CT/MRI image analysis
7. Multilingual support
8. Blockchain-based healthcare records
9. Integration with Large Language Models
10. Real-time patient monitoring systems

These enhancements will improve system scalability, intelligence, and global healthcare accessibility.

16. CONCLUSION

The AI Disease Prediction System provides an intelligent and secure healthcare solution using Machine Learning and Deep Learning technologies.

The system successfully predicts diseases from symptoms and medical images with high accuracy. Ensemble Machine Learning models achieved 89.1% prediction accuracy while the CNN image classification model achieved 91.2% accuracy.

The platform improves healthcare accessibility, reduces diagnosis delays, minimizes manual effort, and supports secure healthcare management.

The project demonstrates how Artificial Intelligence can transform modern healthcare systems through intelligent automation, predictive analytics, and medical image analysis.

Overall, the proposed system provides a practical and scalable foundation for future AI-powered healthcare applications

REFERENCES

[1] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*, MIT Press, 2016.

[2] Tom Mitchell, *Machine Learning*, McGraw-Hill Education, 1997.

[3] TensorFlow Documentation – TensorFlow

[4] Flask Documentation – Flask

[5] OpenCV Documentation – OpenCV

[6] Scikit-learn Documentation – Scikit-learn

[7] Research papers on AI-based healthcare systems and medical image analysis.