

Disease Prediction & Health Analytics System

1. Project Overview

The **Disease Prediction & Health Analytics System** is a data-driven healthcare analytics platform designed to assist hospitals and clinics in predicting diseases, monitoring patient health trends, and identifying abnormal health patterns using statistical and scientific computing.

Modern hospitals generate large volumes of patient data (blood pressure, glucose, cholesterol, heart rate, age, symptoms, etc.), but this data is rarely analyzed for early disease detection. This project bridges that gap by applying **NumPy, Pandas, SciPy, and Matplotlib** to convert raw hospital data into meaningful medical insights.

2. Problem Statement

Most hospitals store patient data in digital files (CSV, Excel, or databases), but:

- They do not perform deep analysis
- Diseases are detected only after symptoms become severe
- Doctors have no visual trend reports
- Risk prediction is done manually

This leads to:

- Late diagnosis
- Increased treatment cost
- Higher patient mortality

3. Proposed Solution

The system will analyze patient medical records and:

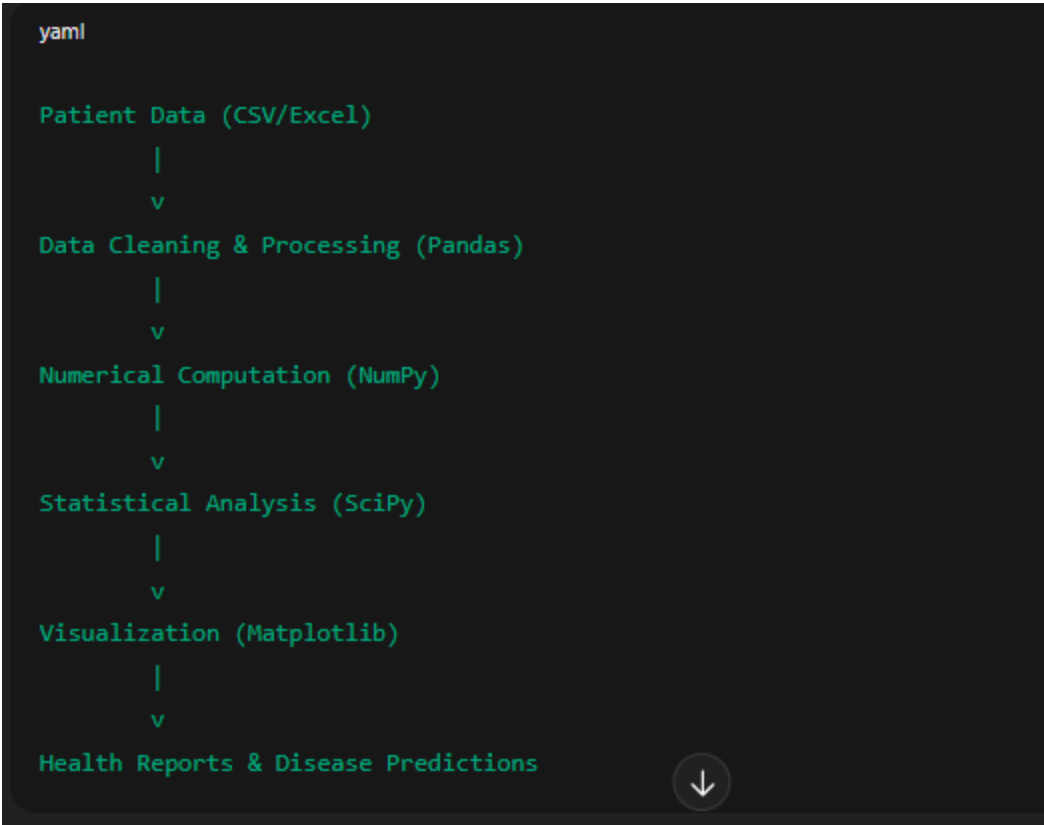
- Predict the probability of diseases such as:
 - Diabetes
 - Heart disease
 - Blood pressure disorders
- Detect abnormal values
- Identify high-risk patients
- Visualize hospital health trends

The system provides doctors with:

- Automated risk scores
- Disease probability

- Interactive medical graphs

4. System Architecture:



5. Dataset Description

Each patient record contains:

| Field | Description |
|----------------|---------------------------|
| Patient_ID | Unique patient identifier |
| Age | Patient age |
| Gender | Male / Female |
| Blood_Pressure | BP level |
| Sugar_Level | Blood glucose |
| Cholesterol | Cholesterol level |

| Field | Description |
|---------------|---------------------------|
| Heart_Rate | Beats per minute |
| BMI | Body mass index |
| Disease_Label | (0 = Normal, 1 = Disease) |

6. Role of Each Library

Pandas

- Load CSV / Excel medical data
- Remove missing values
- Filter patients by age, gender, or disease
- Group by hospital departments

NumPy

- Perform mathematical operations
- Calculate averages, deviations
- Convert datasets into arrays
- Compute health risk scores

SciPy

- Perform statistical tests
- Correlation between BP and heart disease
- Probability of diabetes
- Regression analysis for disease prediction

Matplotlib

- BP vs Age graphs
- Sugar level trends
- Disease frequency charts
- Hospital-wise reports

7. Disease Prediction Model

The system uses **statistical regression and probability models**:

- Logistic Regression (SciPy)

- Correlation Analysis
- Mean, Variance & Standard Deviation
- Risk thresholding

Example:

If

```
Sugar > 180  
BP > 140  
BMI > 30
```

→ High diabetes risk

The system automatically flags such patients.

8. Key Features

Patient Health Analysis

- Identify abnormal BP
- Detect high sugar
- Monitor heart rate

Disease Prediction

- Diabetes risk
- Heart disease risk
- Hypertension risk

Hospital Analytics

- Monthly disease trend
- Age-wise illness distribution
- Gender-wise health patterns

Visualization Dashboard

- Line charts
- Bar graphs
- Risk heatmaps

9. Sample Output

Doctors will see:

- Graph showing diabetes growth over months
- Bar chart of heart disease by age
- Patient-wise risk score
- Alert for high-risk patients

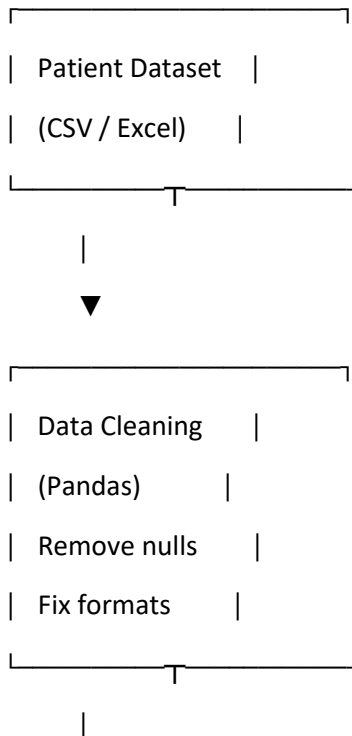
10. Applications

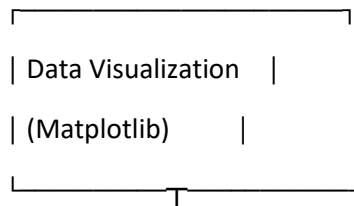
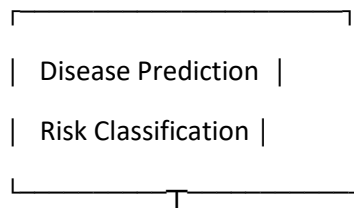
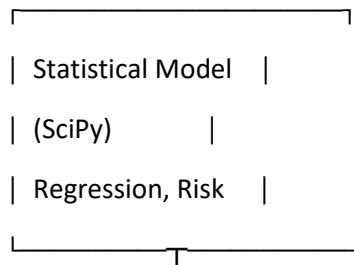
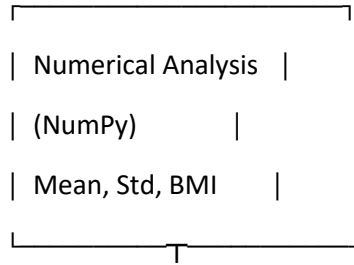
- Hospitals
 - Clinics
 - Health insurance companies
 - Government health departments
 - Medical research institutes
-

11. Future Enhancements

- Machine Learning models
- Real-time patient monitoring
- Integration with hospital databases
- AI-based diagnosis
- Mobile app for doctors

System Flow Diagram:





2. Sample Dataset (CSV Format) You can save this as patients.csv

ID, Age, Gender, BP, Sugar, Cholesterol, HeartRate, BMI, Disease

P001,45,M,150,190,230,88,32,1 P002,32,F,120,95,180,75,22,0 P003,60,M,165,210,260,95,34,1
P004,28,F,110,85,170,70,21,0 P005,50,M,140,160,220,85,29,1 P006,38,F,130,120,200,78,25,0
P007,55,M,155,200,250,92,33,1 P008,40,F,125,110,190,80,24,0

Code Structure:

```
DiseasePredictionSystem/  
|  
├─ data/  
|   └─ patients.csv  
|  
├─ src/  
|   ├─ data_loader.py  
|   ├─ preprocessing.py  
|   ├─ analysis.py  
|   ├─ prediction.py  
|   └─ visualization.py  
|  
├─ main.py  
└─ report/  
    └─ project_report.docx
```

Module Responsibilities

| File | Purpose |
|------------------|----------------------------|
| data_loader.py | Load CSV using Pandas |
| preprocessing.py | Clean and normalize data |
| analysis.py | NumPy & SciPy calculations |
| prediction.py | Disease probability |
| visualization.py | Graphs |
| main.py | System controller |

Chapter 1 – Introduction Background of healthcare analytics

1. Importance of disease prediction
2. Objectives of the project
3. Scope of the system

Chapter 2 – Literature Review Existing hospital systems

1. Traditional diagnosis methods
2. Gaps in current solutions

Chapter 3 – System Analysis Problem definition

1. Feasibility study
2. Functional requirements
3. Non-functional requirements

Chapter 4 – System Design Architecture

1. Data flow diagram
2. Database design
3. Algorithm design

Chapter 5 – Implementation Pandas data processing

1. NumPy calculations
2. SciPy statistical model
3. Matplotlib visualization

Chapter 6 – Results & Discussion Graph outputs

1. Prediction accuracy
2. Health trend analysis

Chapter 7 – Conclusion & Future Work Achievements

1. Limitations
2. Future enhancements

Why this project scores high This project demonstrates:

1. Data Science
2. Healthcare Analytics
3. Statistical Modeling
4. Real-world applicability

It meets:

5. Academic value
6. Industry relevance
7. Research capability