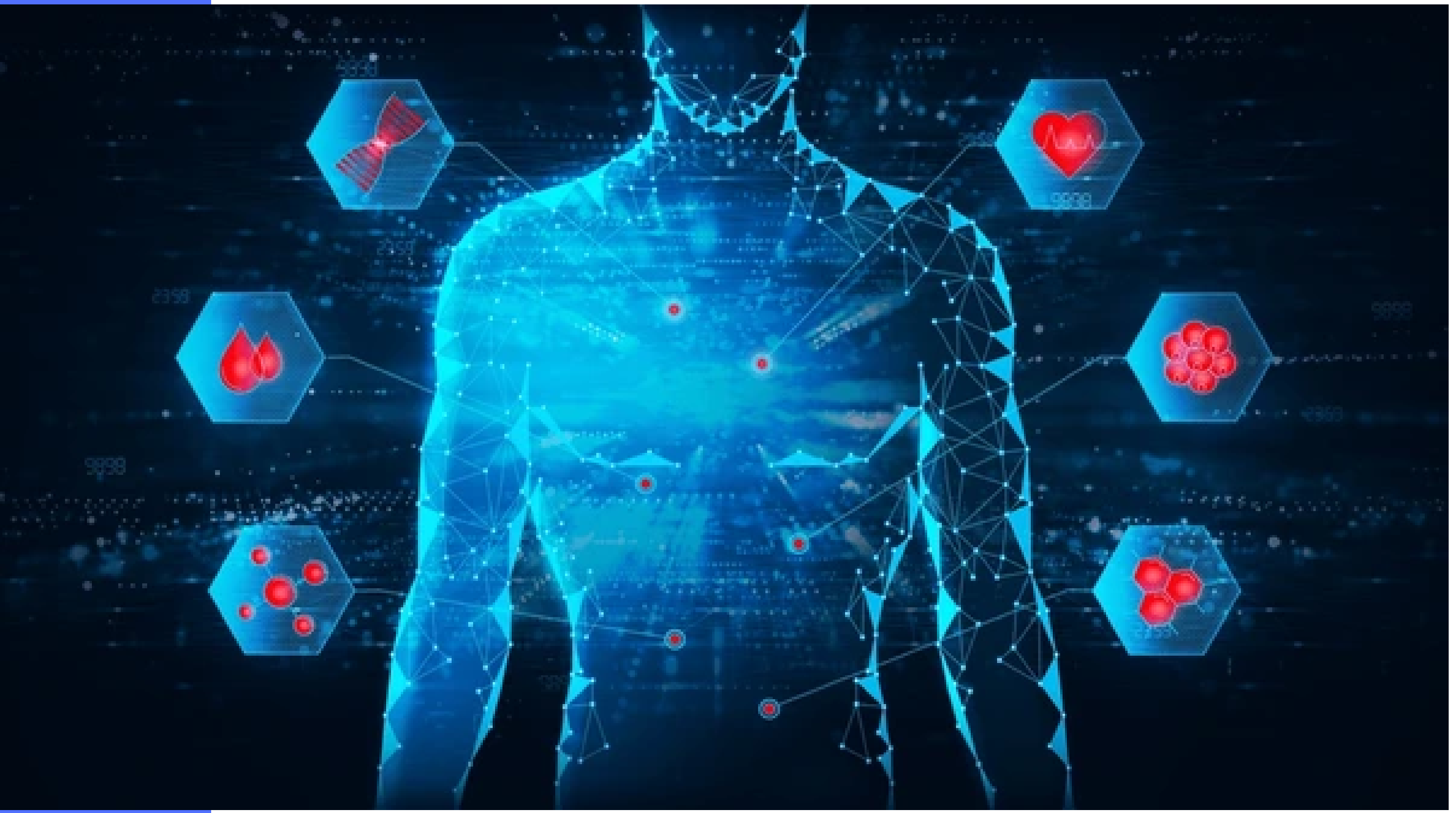
DISEASES PREDICTION USING WEATHER AND SYMPTOMS

Guided By : Abhilasha Joshi Ma’am

Shravan Ghodke 202301070168
Aditya Sonakanalli 202301070175
Samiksha Hubale 202301070178
Aditi Nalawade 202301070179

CONTENT



01 Project Proposal

02 Problem Statement

03 Overview

04 Expected Outcomes

05 Data Loading, Cleaning & Exploration

06 Feature Engineering & Preprocessing

07 Model Training & Evaluation

08 MODEL PREFORMANCE

09 Model Comparison, ROC–AUC & Final Summary

10 Challenges Encountered

11 Deployed Model

PROJECT PROPOSAL

- **Aim:**
To develop a machine learning model that predicts different diseases using weather conditions and symptom data.

- **Dataset:**
Weather-related disease dataset containing:
Temperature, Humidity, Wind Speed, Age, Gender, Symptom indicators, and Prognosis.

- **Method:**
Data cleaning, feature engineering, preprocessing pipeline, and training multiple ML models (Logistic Regression, Random Forest, XGBoost/LightGBM).

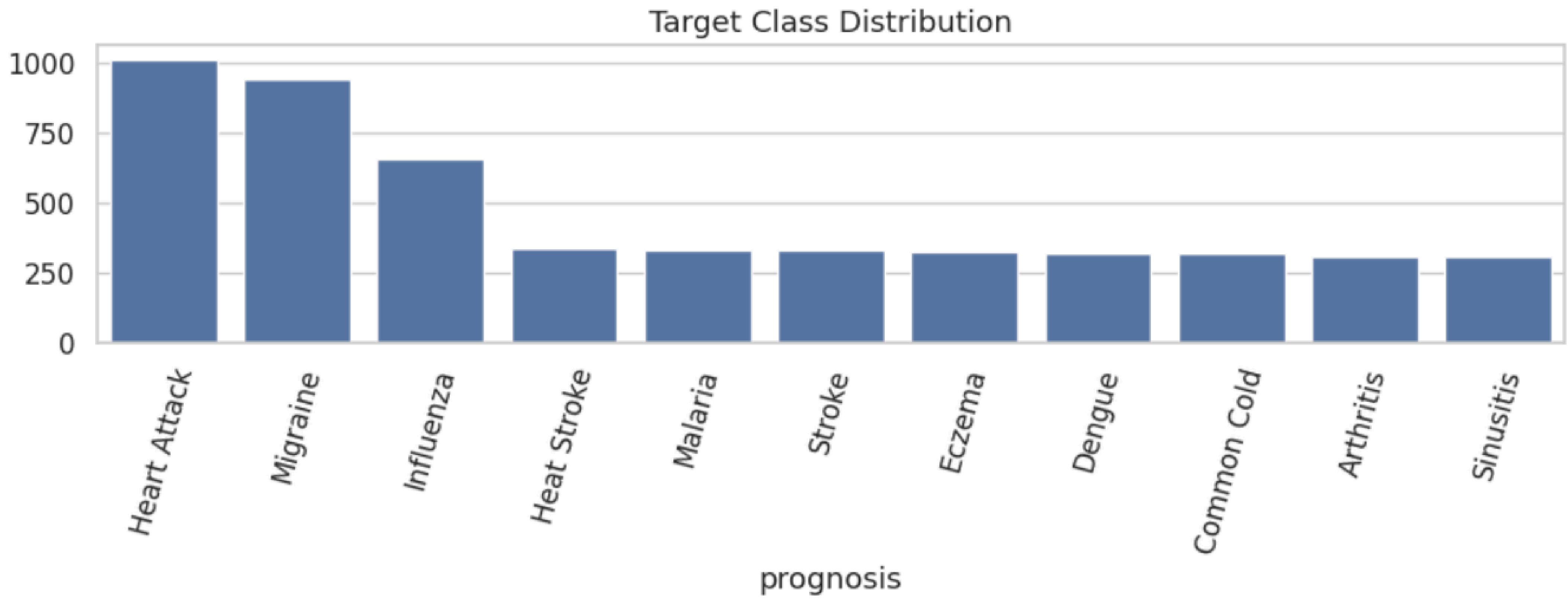
- **Goal:**
To build an accurate and automated disease prediction system that supports early diagnosis and better healthcare decisions.

PROBLEM STATEMENT

- Weather conditions significantly influence the spread of many diseases.
- Changes in temperature, humidity, and wind speed can increase or reduce disease risk.
- This project aims to build a machine learning model that predicts disease prognosis using:
- Temperature
- Humidity
- Wind Speed
- Symptom data
- Demographic information (Age, Gender)
- The system helps in early diagnosis, better health monitoring, and improved medical decision-making.

OVERVIEW

This project predicts the likelihood of different diseases using weather conditions, symptoms, and basic personal details. By analyzing factors like temperature, humidity, and wind speed along with reported symptoms, the machine learning model identifies the most probable disease outcome. The system helps in early diagnosis, better health planning, and timely preventive actions.



EXPECTED OUTCOMES

- A trained ML model that predicts the correct disease using weather and symptom inputs.
- Improved accuracy by combining environmental + symptom features.
- Visualized results using confusion matrices and ROC curves.
- Identification of the best-performing model (Random Forest).
- A deployable model for future integration into apps/web systems.

DATA LOADING, CLEANING & EXPLORATION

Feature Grouping

Weather Features: Temperature, Humidity, Wind Speed

Demographic Features: Age, Gender

Symptom Features: Multiple binary symptom columns

Target Feature: Prognosis (disease label)

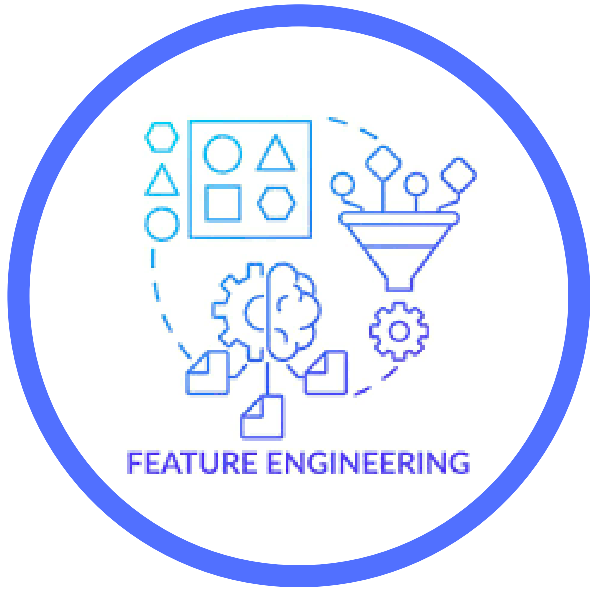
Handling Missing Data

- Symptoms → filled with 0
- Weather + Age → filled with median
- Gender → filled with mode
- Ensures complete and consistent data.

Exploratory Data Analysis (EDA)

- Plotted target class distribution.
- Checked correlations between weather variables and symptom severity.
- Helped understand patterns in data.

FEATURE ENGINEERING & PREPROCESSING

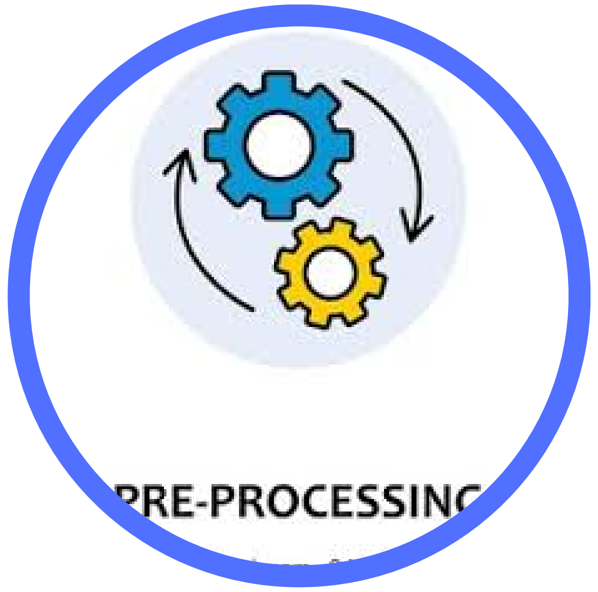


Feature Engineering

- Created new meaningful features:
- symptom_sum: total number of symptoms
- Improves model’s understanding of symptom severity.

Train–Test Split

- 80% training, 20% testing
- Stratified split to keep class distribution balanced
- Target label encoded using LabelEncoder



Preprocessing Pipeline

- Numeric Features: median imputation + scaling
- Categorical Feature (Gender): mode imputation + OneHotEncoding
- Symptom Features: passed without changes
- Built complete preprocessing pipeline using ColumnTransformer and Pipeline.
- Ensures consistent and automated data preparation.

MODEL TRAINING & EVALUATION

Models Trained

- Logistic Regression
- Random Forest
- XGBoost (if available)
- LightGBM (if available)
- Each model combined with preprocessing pipeline.

Training Process

- Fitted pipeline on training data.
- Predicted outcomes on test data.
- Generated classification report for accuracy, precision, recall, F1-score.

Confusion Matrices

- Plotted confusion matrix for each model.
- Helps visualize misclassifications for different disease classes.

Model Performance Storage

- Accuracy and Macro-F1 stored for all models.
- Results displayed in a comparison table.

MODEL PREFORMANCE

MODEL	ACCURACY	MACRO F1
RandomForest	0.990385	0.990531
XGBoost	0.985577	0.983030
LightGBM	0.985577	0.983357
LogisticRegression	0.973077	0.972542

MODEL COMPARISON, ROC–AUC & FINAL SUMMARY

ROC–AUC Analysis

- Used multi-class ROC–AUC to measure how well each model distinguishes between disease classes.
- Models with probability output scored higher.

Performance Comparison Chart

- Bar chart comparing:
  Accuracy
  Macro F1-score
- Visual analysis helped select the best model.

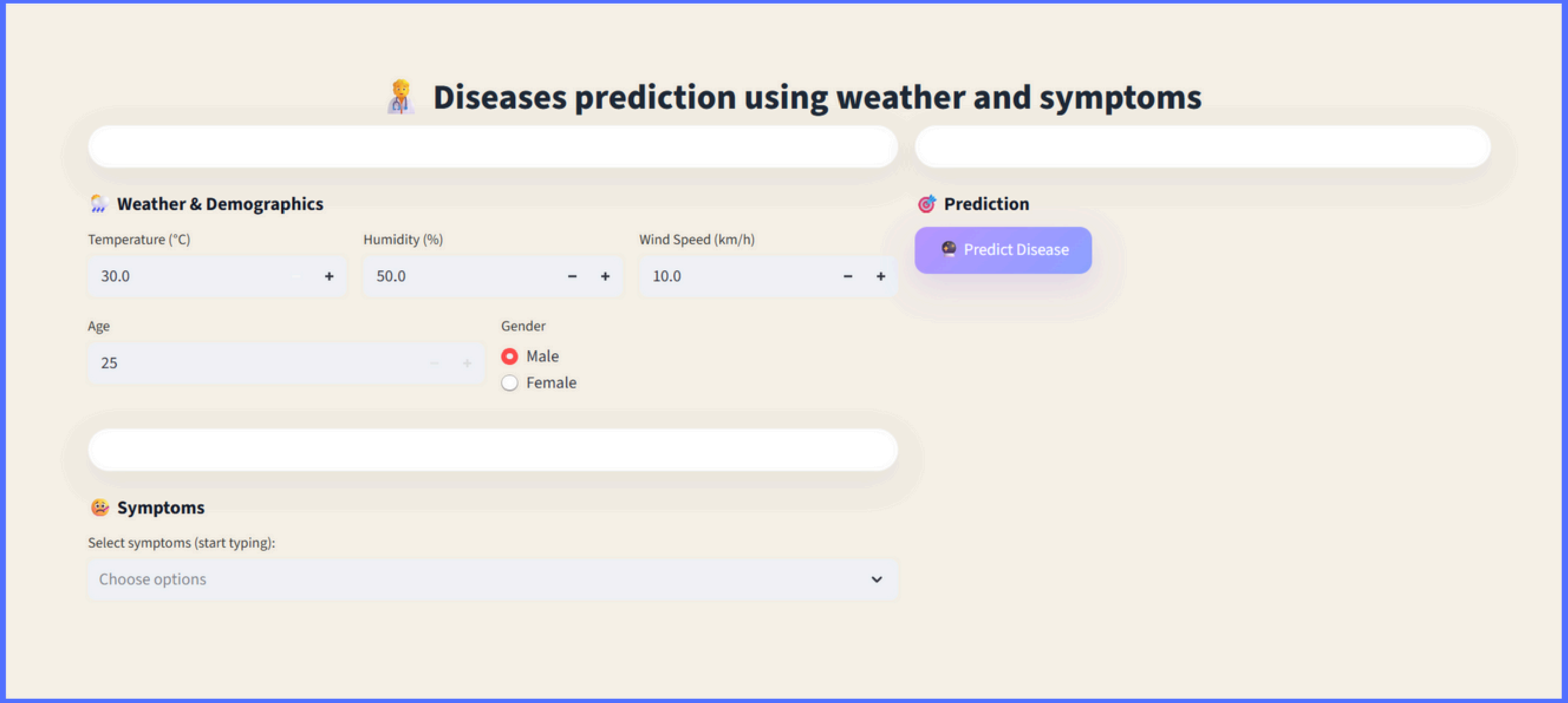
Best Model Selection

- Random Forest achieved the highest accuracy and macro F1-score.
- Stable and effective for multi-class classification.

Final Summary

- Weather features + symptom interactions strongly influence disease prediction.
- Preprocessing pipeline ensures clean and repeatable workflow.
- Random Forest is the best choice for this dataset.

DEPLOYED MODEL



Diseases prediction using weather and symptoms

Weather & Demographics

Temperature (°C): 30.0

Humidity (%): 50.0

Wind Speed (km/h): 10.0

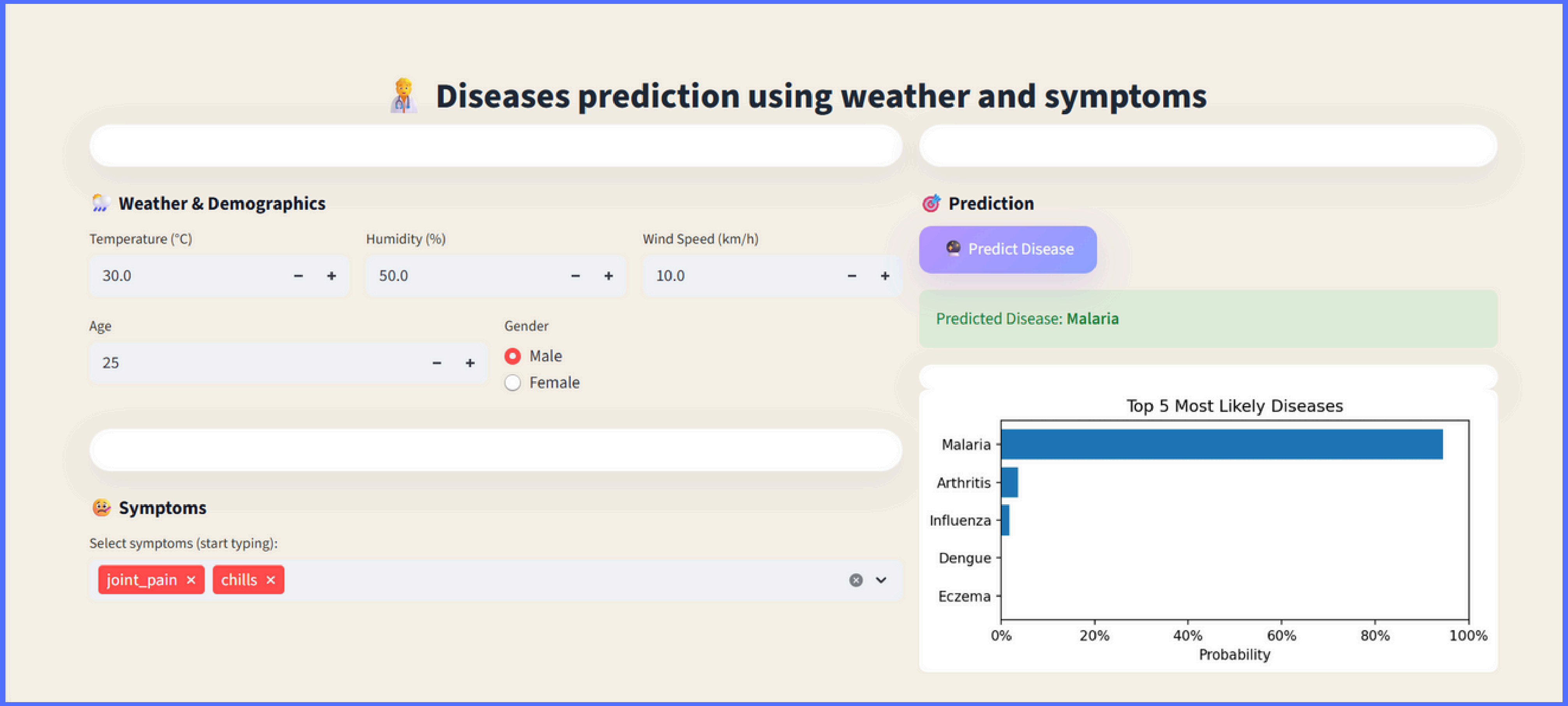
Age: 25

Gender: ● Male ○ Female

Symptoms

Select symptoms (start typing): Choose options

Prediction



Diseases prediction using weather and symptoms

Weather & Demographics

Temperature (°C): 30.0

Humidity (%): 50.0

Wind Speed (km/h): 10.0

Age: 25

Gender: ● Male ○ Female

Symptoms

Select symptoms (start typing): joint_pain × chills ×

Prediction

Predicted Disease: **Malaria**

CHALLENGES ENCOUNTERED

- Dataset imbalance for some diseases.
- Many symptom columns → high dimensionality.
- Cleaning inconsistent weather values.
- Choosing correct preprocessing steps.
- Interpreting multi-class ROC–AUC scores.
- Tuning models to avoid overfitting.

CONCLUSION

- Weather data significantly influences disease prediction.
- Combining symptoms + weather features improves accuracy.
- Random Forest gives the most reliable predictions.
- Useful for early detection and preventive healthcare.

FUTURE SCOPE

- Real-time prediction using IoT weather sensors
- Deploy model via web/mobile app
- More weather parameters (rainfall, AQI, UV index)
- Integration with hospital/clinic databases

THANK YOU