



Engineering
& Computing

Knight Foundation School of Computing and Information Sciences

Fall 2025 Senior Design Project



AI Medical Diagnosis Assistant

Students: Gavin Diaz, Pooja Lad, Stephanie Painchault, AnnCarolyn Power, Rebeca Serralta
Instructor: Seyed Masoud Sadjadi, Florida International University

PROBLEM

Diagnosing genetic diseases often requires multiple tests, manual interpretation of symptoms, and specialized genetic expertise. Patients may present **incomplete or vague symptoms**, and clinicians must compare thousands of possible conditions. This leads to delays, higher costs, and limited access to accurate diagnosis tools.

Our project addresses this challenge by developing an **AI-assisted diagnostic system** that analyzes **patient-reported symptoms** and optional **VCF-encoded genomic variants** to efficiently predict the top probable conditions.

CURRENT SYSTEM

Current diagnosis depends on clinician's experience, manual review of symptoms, and genetic reports. Existing tools require structured medical data and cannot easily handle flexible or incomplete symptom input. Presently, no accessible system accepts both symptoms and optional VCF files to generate AI-based predictions.

VERIFICATION

Test Case 1: Symptom-Only Input

- Input:** Symptoms (e.g., weight loss, pruritus)
- Expected:** Low confidence due to non-specific symptoms
- Result:** System spreads confidence across 5 probable conditions

Test Case 2: Symptoms + VCF Input

- Input:** Symptoms (e.g., weight loss) + pathogenic variants (VCF)
- Expected:** Higher confidence and clearer condition mapping
- Result:** System identifies top 5 likely conditions with increased confidence

Test Case 3: Healthy Control

- Input:** No symptoms + benign VCF
- Expected:** Classified as “healthy” or very low confidence across diseases
- Result:** Model outputs near-flat low confidence

IMPLEMENTATION

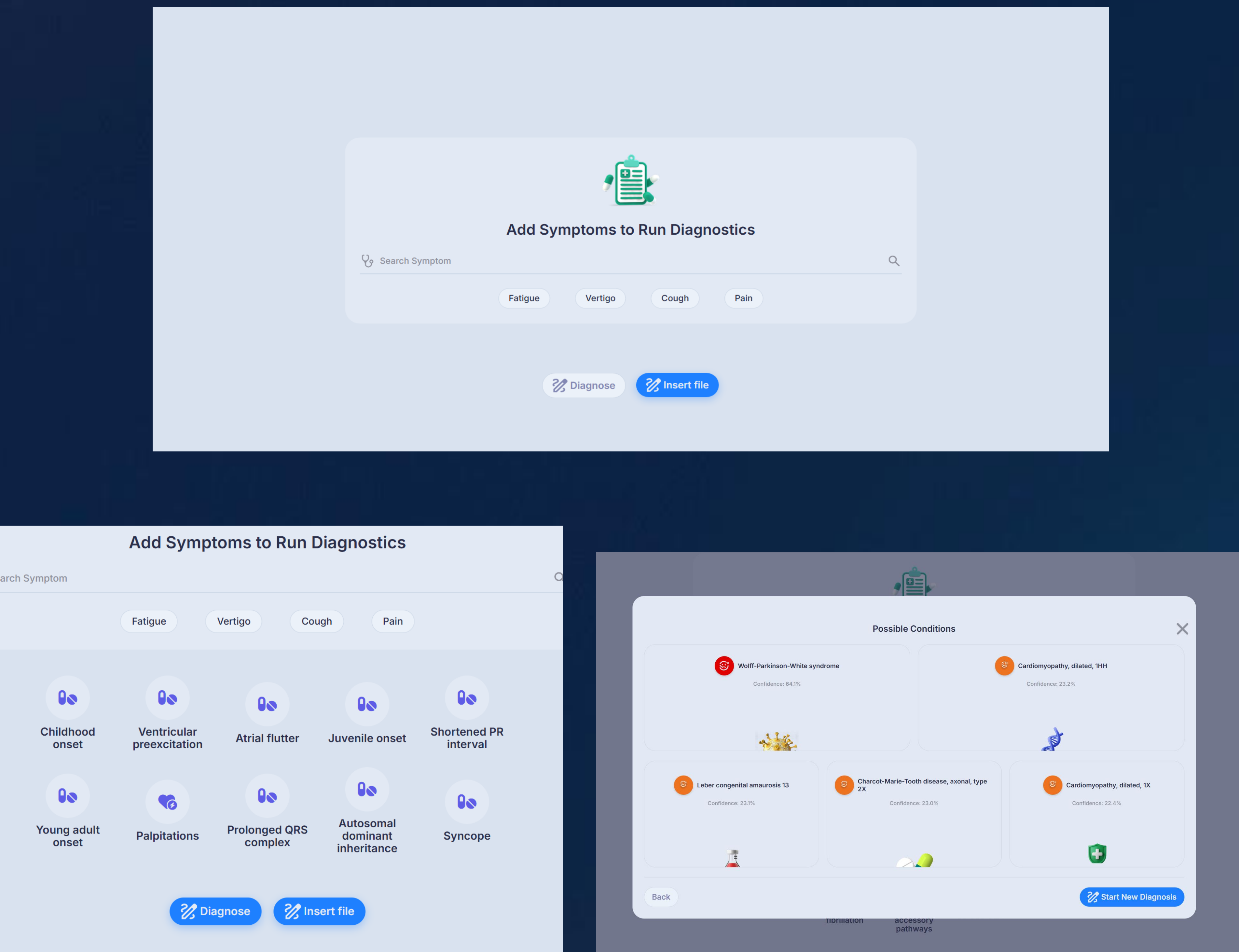


SUMMARY

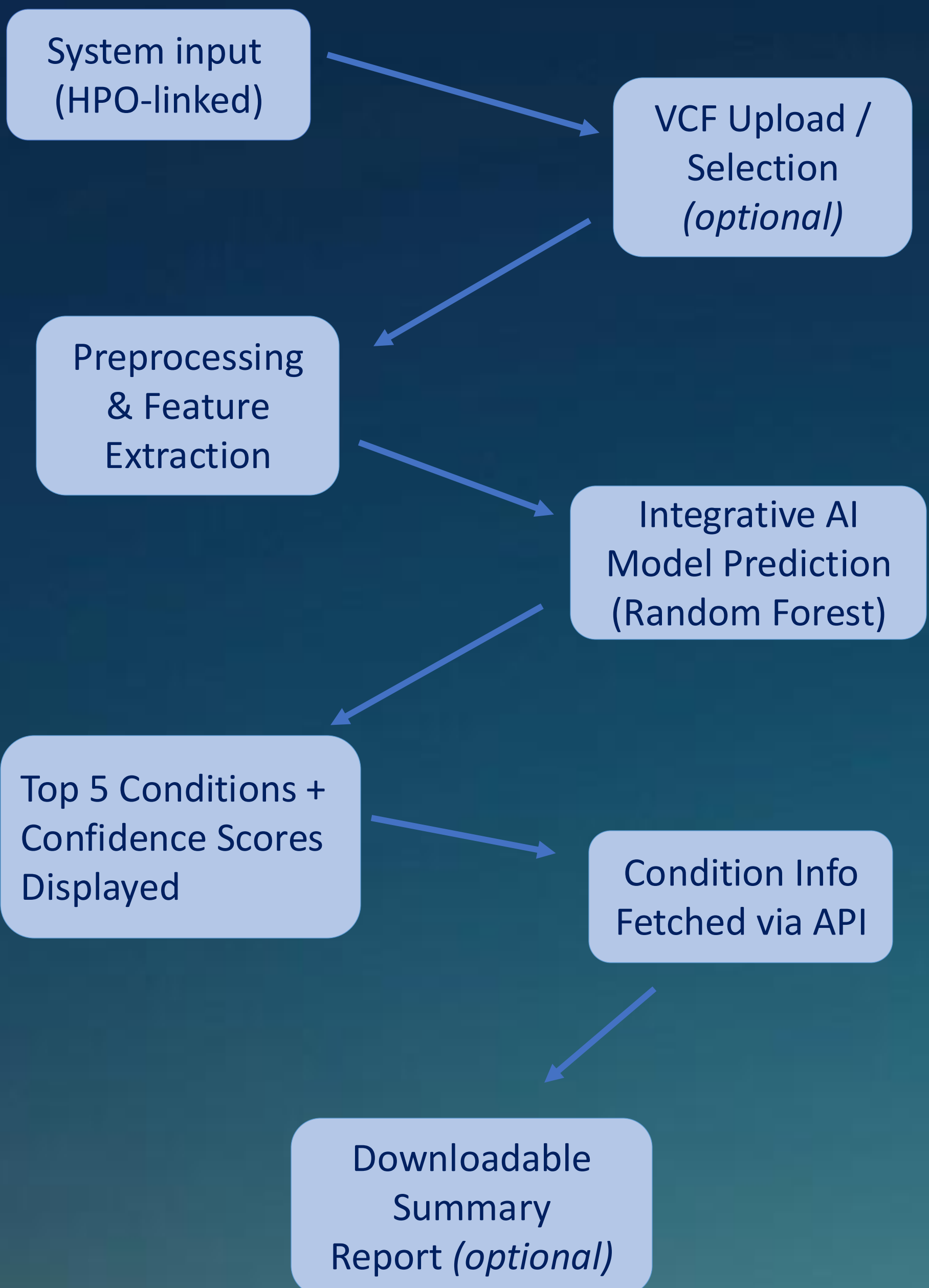
Our project is an **AI-powered diagnostic assistant** designed to analyze patient **symptoms** and optional **genomic variant data** (VCF files). Users enter any number of HPO-linked symptom terms and may select a sample encoded VCF file. The backend processes these inputs, maps symptoms to their corresponding HPO IDs, extracts key variant information, and evaluates them using an **integrated prediction pipeline** that combines a trained Random Forest model, phenotype-condition similarity scoring, and (when available) variant-level associations.

The system outputs the **top five most likely conditions** with confidence scores, and, when available, supplements each result with concise medical context drawn from public clinical data sources. The goal is to create a decision-support tool that accelerates and verifies early identification of likely diagnoses, especially when symptoms are vague, incomplete, or clinically ambiguous.

OBJECT DESIGN



SYSTEM DESIGN



Functional requirements

REQUIREMENTS

- Accept user-entered symptom inputs (HPO terms)
- Optionally select a sample VCF file
- Run the integrative model and output the top 5 predicted conditions
- Display prediction confidences and pull additional condition information via API
- Provide a web interface allowing flexible symptom input and file upload

Technical requirements

- Model trained on curated dataset consisting of 59 conditions, ~100 patient cases each post-preprocessing
- Each patient case must include: symptoms, a VCF file path, and a classification label
- Machine learning model with validated accuracy and no data leakage
- Backend API connecting frontend inputs → integrated ML model → prediction results
- Secure handling of genomic files and consistent relative directory paths

ACKNOWLEDGEMENT

We thank Dr. Seyed Masoud Sadjadi for his assistance and mentorship that we received throughout the senior design project.