Food-Drug Interaction and Allergy Prediction System Using Machine Learning

G.G.R.M. SAHAN PRABUDDHA

A dissertation submitted in partial fulfilment of the requirement for

Bachelor of Engineering (Honours) degree in Software Engineering

School of Computing

Informatics Institute of Technology, Sri Lanka

in collaboration with

University of Westminster, UK

2025

ABSTRACT

Problem: Individuals with allergies often struggle to identify potential allergic reactions when consuming packaged food or taking multiple medications, especially in cases where ingredient lists are complex or not clearly understood. Existing systems typically focus on either drug-drug interactions or single-allergen predictions, lacking a comprehensive, personalized approach. This research addresses this gap by developing a system that predicts potential allergic reactions caused by interactions between food and medicine ingredients, considering individual-specific health data such as gender and blood type.

Methodology: The system utilizes Optical Character Recognition (OCR) to extract food label text and combines it with user-entered medication details. Two machine learning models are developed—one for predicting allergy severity using Gradient Boosting, and another for predicting allergy descriptions using XGBoost and Random Forest. The dataset is preprocessed using normalization and SMOTE for balancing. Personalized features such as gender and blood group are integrated to improve model accuracy. The system is built as a web-based platform using React.js for the front end and a Python backend for model execution and prediction.

Initial Result: The allergy severity prediction model achieved an accuracy of **96.27%**, while the allergy description prediction model achieved **84.01%** (XGBoost) and **77.52%** (Random Forest). OCR accuracy depends on label clarity and formatting but generally performs well under clean image conditions. Although results show strong initial performance, improvements are planned for label variability handling, ingredient name normalization, and large-scale testing with real-world data.

Subject Descriptors:

- Applied computing → Health Informatics
- Computing methodologies → Machine learning → Supervised learning → Ensemble methods
- Human-centered computing → Accessibility → Health-related applications

Keywords: Allergy prediction, food-drug interaction, machine learning, OCR, personalized health data, severity classification, ensemble models, allergy detection system