Drug-Induced Autoimmunity Prediction Model using Ensemble Machine Learning

Author: Srikar Sistla (Student ID ZE65892)

Project Overview
This research project focuses on developing ensemble machine learning models, specifically using Random Forest (RF), Gradient Boosting (GB) and Support Vector Machine (SVM) classifiers to predict drug-induced autoimmunity using molecular descriptors [1]. We utilize these classifiers to analyze chemical structures and their potential autoimmune effects. Results of this research will be submitted to the Fourth Annual Early Career Symposium at the Institute for Bioscience & Biotechnology Research.

Background
Drug-induced autoimmunity occurs when the body's immune system is triggered by a medicine. Drug-induced autoimmunity represents a significant challenge in pharmaceutical development, with various mechanisms contributing to adverse immune responses [2]. Recent studies have highlighted the importance of understanding molecular features that may predispose drugs to trigger autoimmune reactions [3]. Classifying drugs as potentially causing autoimmunity problems is very complex. It has large domain space size which includes both continuous and discrete values. The dataset being used is imbalanced. And the magnitude of attributes varies greatly. Predicting drug-induced autoimmunity is a very complex problem.

Methodology
As part of this research an Ensemble Machine Learning model will be created combining predictions from three other models: RF, GB and SVM. We will utilize a dataset of **477 drug compounds**, each characterized by **198 molecular descriptors** such as the Balaban J index, Bertz CT, and topological indices. The task is binary classification-distinguishing compounds that induce autoimmunity from those that do not. Features will be standardized using Standard Scaler to ensure consistency across the model inputs.

Mechanistic Insights
Emerging studies have linked drug-induced autoimmunity to certain structural and physicochemical features. These include the drug's ability to bind to **HLA molecules**, its molecular complexity and size, presence of specific functional groups, and electronic properties. Understanding these patterns helps reveal how certain drugs may inadvertently activate the immune system.

Clinical Implications
Predicting autoimmune potentiality in drugs could significantly enhance **drug development practices** by flagging problematic compounds early. It also enables patient-specific risk assessments, supports better monitoring during treatment, and aids in designing preventive strategies to minimize adverse effects.

Expected Impact
This study aims to improve early detection of autoimmunity-inducing drugs, potentially reducing development costs and enhancing safety assessments. Moreover, it seeks to provide deeper insights into how specific molecular structures correlate with autoimmune risks, helping guide safer drug design. The use of **Ensemble Machine Learning** enhances the robustness and generalizability of these predictions.

Resource Requirements
Successful execution of this project will require computational infrastructure for model training, access to chemical structure databases, and collaboration with experts in medicinal chemistry and toxicology for interpretation and validation of findings. All resources have currently been obtained.

This will be interdisciplinary research combining both Data Science, Biology (Biochemistry) and Pharmaceuticals. I will provide Data Science expertise but will partner with Kylie Carter, UMBC Biology student. Ms. Carter will provide Biology and Pharmaceutical expertise on this project.

References
1. Wu Y, et. al. Machine Learning for Predicting Risk of Drug-Induced Autoimmune Diseases by Structural Alerts and Daily Dose. International Journal of Environmental Research and Public Health. 2021; 18(13):7139.
2. Guo H, et. al. Modeling and insights into the structural characteristics of drug-induced autoimmune diseases. Front. Immunol. 2022; 13:1015409.
3. Huang L, et. al. InterDIA: Interpretable prediction of drug-induced autoimmunity through ensemble machine learning approaches. Toxicology. 2025 Feb; 511:154064.

SRIKAR SISTLA

Halethorpe, MD

📞 443-636-7777 ✉ srikarsistla710@gmail.com in LinkedIn ○ Github 🌐 Portfolio

Professional Summary

Innovative and results-driven Business Intelligence Front-end Engineer with expertise in interactive data dashboards, analytics platforms, and real-time reporting tools. Proficient in Power BI, SQL, Python and React.js, integrating business intelligence with front-end development to create user-friendly, data-driven applications. Passionate about data storytelling, business reporting, and designing interactive intelligence solutions for enterprise users.

Software Engineer specializing in API development, full-stack web applications, and business intelligence solutions using Python3, React.js, SQL, and cloud services (AWS, Firebase).

Proficient in RESTful APIs, GraphQL, database operations (CRUD), ETL pipelines, and CI/CD automation for scalable applications.

Experience in machine learning, predictive analytics, and web scraping to enhance data-driven decision-making.

Experience

TANTVSTUDIOS   **Nov 2024 – Present**
Software Engineer Intern   *Washington DC, DC*
- Developed an interactive **Website for modern NEWS** using **React.js, Next.js and typescript**.
- Integrated **REST APIs & GraphQL** for real-time data exchange, increasing performance by 40%.
- Integrated crucial middleware and backend elements using **C#**
- Implemented **PostgreSQL CRUD operations** for efficient data retrieval and updates.
- Optimized page load speed, improving frontend performance by **30%.**

Valhalla Data Systems   **Oct 2024 – Nov 2024**
Front End Developer Intern   *Clarksburg, MD*
- Designed and launched a Full Stack educational website using **PHP and MySQL**,
- Launched and publsihed **Olademy, enhancing Maryland High School education access and enabling free education from top-tier universities.**
- Revamped design tools and processes, leading to an **80%** increase in annual revenue.
- Utilized **PHP, HTML, CSS, and Linux OS** for seamless integration and development.

Santa's Knights   **Aug 2024 – Oct 2024**
IT Project Manager   *New York City, NY*
- Led the project "Equisym" as **Project Manager** by applying business operation such as **Agile Methodology, Kanban Boards** and various other principles for a successful launch.
- Managed cross-functional teams, reducing project delays by 25% through improved collaboration, Implemented unique PM tools such as **JIRA, CLICKUP** for timely deliverables.
- Used **Power BI and Excel** to create business intelligence reports for growth analysis.

Expand AI   **Sept 2021 – Feb 2022**
Data Analyst Intern   *Bengaluru, KT, India*
- Developed an **ETL pipeline** for data preprocessing and transformation using **SQL and Python**.
- Built an interactive **Power BI dashboard**, integrating with a **Flask API** for real-time business insights.
- Implemented **web scraping scripts (BeautifulSoup, Scrapy)** to extract external business intelligence data.
- Automated **data extraction & storage** using **AWS Lambda & RDS**.
- Enhanced **CI/CD pipeline** for API deployments using **GitHub Actions & Jenkins**.

Education

University of Maryland at Baltimore County   **Sept 2023 – Dec 2025**
Master of Science in Information System   *Baltimore County, Maryland*

SCSVMV University   **Aug 2019 – June 2023**
Bachelor of Engineering in Computer Science   *Kanchipuram, Tamil Nadu, India*

Technical Skills

Programming Languages: Python3, JavaScript, TypeScript, SQL (PostgreSQL, MySQL), Bash

Frontend Development: React.js, Next.js, HTML5, CSS3, Redux, Apollo

Backend & API Development: Node.js, Express.js, Flask, FastAPI, REST APIs, GraphQL

Database Management: PostgreSQL, MySQL, MongoDB (CRUD operations)

Cloud & DevOps: AWS (S3, Lambda, RDS), Docker, Kubernetes, Jenkins CI/CD, GitHub Actions

ETL & Data Processing: Apache Airflow, dbt (Data Build Tool), Web Scraping (BeautifulSoup, Scrapy)

Projects And Publications

Prediction of Diabetes using SVM | *Python, ML Algoritms*   **May 2022**
- Developed an ML-based classification model using Support Vector Machines (SVM) to predict diabetes risk, achieving 80.5% accuracy
- Built interactive Power BI dashboards for medical professionals, enabling real-time visual insights into patient data
- Implemented data cleaning, feature selection, and model evaluation using Python (Pandas, NumPy, Scikit-learn)
- Published findings in IJSCE Journal (DOI: 10.35940/ijsce.B3557.0512222)

Case Study on Revenue Analytics | *Data Visualization, Information Processing, POWER BI*   **Nov 2024**
- Conducted Revenue & Profitability Analysis using Power BI, SQL, and Python, identifying trends and revenue drivers for business growth.
- Built machine learning models in Python (Scikit-learn, Pandas) to forecast future revenue trends based on historical data.
- Designed custom Power BI reports & dashboards, providing data-driven insights for stakeholders & decision-makers.
- Automated ETL pipelines & API integrations for real-time data updates, improving reporting efficiency by 30%.

A Genetic Algorithm for Rule Generation to Predict Student Success | *Machine Learning, Predictive Analytics***March 202**
- Developed a **Genetic Algorithm**-based model to predict student success using **rule-based classification**.
- Performed **feature selection, data preprocessing**, and model optimization on **UC Irvine dataset**.
- Implemented **Python (Scikit-learn, Pandas)**, improving prediction accuracy to **24.2 fitness score**.
- Published in academic research (View Publication).