Bioinformatics Spring 2025

Precision Medicine Using Genomics

Mid Report

Section: 8B

Group Members:

Muhammad Laraib Akhtar 21L-5294

Wajeeh Ahmed Khan Niazi 21L-7565

Syed Ahmer Zaidi 21L-5284

National University Of Computer and Emerging Sciences
Department of Computer Science
Lahore, Pakistan

1. Project Overview

The project titled **"Precision Medicine Using Genomics"** focuses on developing a machine learning-based system for drug response prediction. The primary objective is to personalize treatment plans based on a patient's genetic profile, minimizing adverse drug reactions and maximizing therapeutic efficacy. The project centers around predicting patient responses to medications such as Warfarin and Metformin using pharmacogenomic data.

The timeline encompasses tasks from initial dataset exploration to final model deployment, aiming for a complete ML-powered web solution. The key stakeholders include project group members **Wajeeh Ahmed**, **Laraib Akhtar** and **Ahmer Zaidi,** along with potential end-users like clinicians and patients who will benefit from the tool's implementation.

2. Progress Summary

Major Milestones Achieved:

- **Drug Selection**: Warfarin and Metformin were shortlisted based on genetic relevance and dataset availability.
- **Dataset Preprocessing**: Leveraged datasets from PharmGKB and the UK Biobank; handled missing values, encoded genetic variables, and normalized features.
- **Model Training**:
  - *Warfarin Dose Prediction*: Models like Random Forest and XGBoost outperformed baselines with a Mean Absolute Error of ~6 mg/week.
  - *Metformin Response Classification*: Achieved ~80% accuracy with Random Forest/XGBoost models.
- **Backend Development**: Initiated with FastAPI to serve ML predictions.
- **Frontend Development**: UI wireframes created using Next.js, along with an input form for clinical and genomic data.

Adjustments to Scope or Timeline: There have been no significant deviations in scope; however, model complexity and real-time deployment concerns have adjusted some backend and frontend timelines.

3. Achievements and What Worked Well

- **Effective Dataset Curation**: Choosing well-documented and annotated datasets from reputable sources (e.g., PharmGKB) streamlined preprocessing and modeling.
- **Model Performance**: Random Forest and XGBoost consistently delivered strong performance in both regression and classification tasks.
- **Technology Stack**:
  - *FastAPI* enabled rapid backend development and easy integration with ML models.
  - *Next.js* proved suitable for building a responsive frontend.

- **Clear Workflow Planning**: Establishing a phased approach (dataset → model → backend → frontend) helped maintain structured progress.

These components collectively contributed to maintaining alignment with the original plan while ensuring consistent progress.

4. Challenges and What Did Not Work

- **Data Complexity**: Missing values and categorical genetic markers posed difficulties during preprocessing.
    - *Resolution*: Used statistical imputation and proper encoding strategies.
- **Model Interpretability**: Clinicians require clear explanations for predictions, which typical ML models do not offer.
    - *Resolution*: Introduced SHAP for interpretability, allowing genetic feature importance visualization.
- **Deployment Complexity**: Building an efficient, user-friendly web app required careful coordination between backend and frontend.
    - *Resolution*: Chose a modular architecture (FastAPI + Next.js) to streamline integration.

Lessons Learned:

- Preprocessing genomic data requires domain understanding.
- Interpretability is crucial in medical applications, and must be integrated from early stages.
- Early selection of a scalable stack helps avoid later reworks.

5. Future Plans

Key Tasks Remaining:

- **Model Optimization**:
    - Hyperparameter tuning of existing models.
    - Adding more clinical features to improve predictions.
    - Finalize SHAP/LIME implementation.
- **Backend Finalization**:
    - Complete API deployment.
    - Improve speed and reliability of real-time predictions.
- **Frontend Completion**:
    - Connect form inputs to backend.
    - Implement visual result feedback and add session/user management.
- **Final Testing & Documentation**:
    - Evaluate using real-world test cases.

  - Compare outputs with clinical guidelines.
  - Compile final report and research paper.

Resource Needs:

- Access to a secure server for API deployment.
- Additional annotated patient data for robust validation.

6. Updated Expectations and Outcomes

The core goals of the project remain unchanged; however, there is now a stronger emphasis on **explainability and user experience**. Based on current findings:

- **Expected Outcomes** now include a **clinician-friendly interface** and **model transparency via SHAP**.
- **Scope Adjustments**: Slight broadening to include additional variables and interpretability tools.

The deliverables are still in alignment with the original vision but are now enhanced by the incorporation of medical AI best practices.

References

[1] M. Wadelius et al., "Estimation of the Warfarin Dose with Clinical and Pharmacogenetic Data," *New England Journal of Medicine*, vol. 360, no. 8, pp. 753–764, Feb. 2009. [Online]. Available: https://www.nejm.org/doi/full/10.1056/NEJMoa0809329

[2] M. S. Lee et al., "Effects of putative metformin targets on phenotypic age and leukocyte telomere length: a Mendelian randomisation study using data from the UK Biobank," *The Lancet Healthy Longevity*, vol. 4, no. 2, pp. e85–e94, Feb. 2023. [Online]. Available: https://www.thelancet.com/journals/lanhl/article/PIIS2666-7568(23)00085-5/fulltext

[3] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*, 2017. [Online]. Available: https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[4] S. Ramírez, *FastAPI*. [Online]. Available: https://fastapi.tiangolo.com/

[5] Vercel, *Next.js Documentation*. [Online]. Available: https://nextjs.org/docs