Machine Learning Applications in Noninvasive Prenatal Testing and Carrier Screening

Yasmin Rocio Orduz Landazabal
Department of Computer Science
Georgia Southern University
Statesboro, GA, USA
yo00553@georgiasouthern.edu

***Abstract*—This paper explores the application of machine learning and data science techniques in Noninvasive Prenatal Testing (NIPT) and carrier screening. NIPT uses cell-free fetal DNA from maternal blood to detect chromosomal abnormalities without invasive procedures. We investigate how advanced algorithms can improve accuracy, reduce false positives, and enhance early detection of genetic conditions. This research aims to demonstrate the potential of AI-driven approaches in revolutionizing prenatal healthcare and genetic screening.**

***Index Terms*—Noninvasive Prenatal Testing, NIPT, Machine Learning, Genetic Screening, Carrier Screening, Healthcare AI, Chromosomal Abnormalities**

I. Project Repository

All project materials, code, and documentation are available in our GitHub repository at: https://github.com/Yasmin-maker1/DS_ML_Project_colab-integration

II. Introduction

Noninvasive Prenatal Testing (NIPT) represents a breakthrough in prenatal care, allowing detection of chromosomal abnormalities through analysis of cell-free fetal DNA in maternal blood. Traditional screening methods often require invasive procedures with associated risks. Machine learning algorithms can analyze complex genomic data patterns to improve diagnostic accuracy and identify potential genetic conditions earlier in pregnancy. This project will explore various ML techniques including classification algorithms, neural networks, and statistical modeling to enhance NIPT accuracy and carrier screening effectiveness.

III. Literature Review

We will review existing research on machine learning applications in genetic testing, focusing on deep learning approaches for DNA sequence analysis and pattern recognition in prenatal screening data. Recent studies have shown promising results in using convolutional neural networks for genomic sequence classification.

IV. Methodology

Our approach will involve collecting publicly available genomic datasets, preprocessing the data through normalization and feature extraction, implementing various machine learning models including Random Forest, Support Vector Machines, and Deep Neural Networks, and evaluating their performance in detecting chromosomal abnormalities such as trisomy 21, 18, and 13.

V. Expected Outcomes

We anticipate demonstrating improved accuracy rates in genetic condition detection compared to traditional statistical methods, and developing a framework for integrating ML models into clinical NIPT workflows while maintaining interpretability for healthcare professionals.

VI. Data Sources

We plan to utilize publicly available datasets from repositories such as NCBI GenBank, dbGaP (Database of Genotypes and Phenotypes), and published research studies on prenatal genetic screening. All data will be anonymized and used in compliance with ethical guidelines.

VII. Ethical Considerations

This research will address privacy concerns related to genetic data, data security in cloud-based ML systems, potential biases in training data, and ethical implications of AI-driven genetic testing in healthcare decision-making.

VIII. Conclusion

This project aims to contribute to the advancement of prenatal healthcare by leveraging machine learning to make genetic screening more accurate, accessible, and reliable. The integration of AI technologies in NIPT has the potential to reduce false positives, improve early detection rates, and provide expecting parents with more reliable information for informed decision-making.

References

[1] Author Name, "Title of Paper," *Journal Name*, vol. X, no. Y, pp. 1-10, 2024.
[2] Author Name, "Another Paper Title," *Conference Name*, pp. 1-5, 2024.