**ABSTRACT**

Down Syndrome (DS), or trisomy 21, is a congenital genetic disorder characterized by the presence of an extra chromosome 21. It leads to developmental delays, intellectual disabilities, and various health complications. Traditional diagnostic methods, including karyotyping and non invasive prenatal testing (NIPT), while effective, are either time-consuming, expensive, or inaccessible in resource-limited settings. With the advent of high-throughput technologies such as microarrays, vast gene expression data can be leveraged to develop computational models for rapid, scalable, and cost-effective diagnosis. This project proposes an end-to-end machine learning pipeline for the prediction of Down Syndrome using gene expression profiles sourced from public datasets (GSE6408 and GSE9321) in the Gene Expression Omnibus (GEO). Raw .gpr files were processed to extract meaningful features such as normalized signal intensities (e.g., CH1 and CH2 signal and background medians), with poor-quality spots filtered based on flag values. The pipeline incorporates key preprocessing steps, including missing value handling, synthetic minority oversampling (SMOTE) for class imbalance, noise injection for robustness, and standardization using StandardScaler. Several machine learning models—Random Forest, XGBoost, and LightGBM—were trained and evaluated for classification accuracy. The Random Forest model demonstrated superior performance and was deployed alongside the scaler in a web application using Flask. The app enables users to upload microarray-derived CSV files, automatically processes the data, and returns predictions on whether each sample indicates the presence of Down Syndrome. This project highlights the feasibility and effectiveness of integrating genomic data analysis with machine learning for the early detection of genetic disorders. It underscores the potential of AI driven biomedical tools in aiding clinical decisions, particularly in genomics-based diagnostics