INFORMATICS INSTITUTE OF TECHNOLOGY

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

Prediction Of Rare Genetic Diseases Using Facial Images

A Project Proposal by

Miss. Baveega Kuganesan

Supervised by

Mr. Rathesan Sivagnanalingam

Submitted in partial fulfilment of the requirements for the BEng (Hons) Software Engineering Degree at the University of Westminster.

May 2023

ABSTRACT

Medical research has a major problem because of the genetic data's explosive increase, especially when it comes to the early identification and identification of genetic illnesses. Facial images is one potentially significant but underused source of data, as some genetic illnesses frequently produce distinctive and recognizable facial traits. Traditional approaches of picture analysis, however, take a long time, demand a high level of knowledge, and are subject to human mistake. Furthermore, a more effective and precise technique is required due to the enormous number of genetic data. An automatic, accurate, and scalable method to recognize facial phenotypic characteristics linked to genetic illnesses is urgently needed.

Created an advanced machine learning model to solve this issue utilizing deep convolutional neural networks (CNNs), specifically the ResNet50 and VGGFace algorithms. Because they can extract intricate features from simple pixel input, CNNs are particularly well-suited for image analysis. ResNet50 was used because of its outstanding performance in picture classification tasks and because of its deep residual learning framework, which helped to solve the disappearing gradient issue. This was paired with VGGFace, a model that had already been trained and was renowned for being good at facial recognition tasks. A sizable collection of facial photos with associated genetic abnormalities was used to train the model. The architecture was created to employ the ResNet50 model to categorize the genetic condition after passing these focused images via the VGGFace models to recognize and fix on the facial areas in the photos.

Accuracy, precision, recall, and F1 score are common data analysis metrics for classification tasks that were used to assess the performance of our model. This was accomplished by using a different test dataset that the model has never seen before during training. The model displayed good accuracy, showing a high percentage of overall accurate predictions. A low percentage of false positives, which is crucial in a medical environment to prevent needless treatment or discomfort, was also indicated by excellent precision. The model's capacity to accurately identify the majority of cases of genetic illnesses was shown by the recall score, which was equally outstanding. The F1 score, which balances precision and recall, supported our model's strong performance. This cutting-edge method for identifying genetic disorders offers a promising course for ongoing study and therapeutic applications.