Abstract:

Reported rare diseases are over 8000 around the world .while each disease is different from another. Estimation say that 1 between 15 or 20 person has some rare diseases. Mostly they are caused because of 1 or 2 changes happened in the genome. The causative variant need to be identified from the others that make a distinction between one person's genome from another, that is not an easy task to be done. In this article, we describe an overview of the process that happened on data through multi-stage of diagnosing process. In every stage, we discuss algorithms and methods used in diagnostic laboratories and how we enhancing the process in general with machine learning and with deep learning specifically.

Introduction:

In 2000 was the first released of a draft human genome with covering percent ~95, commonly referred as human genome reference, sequence created by sequencing and assembling DNA collected by unknown volunteers. This ~3 billion nucleotide long genome has several revision over years also there are a few small regions that have remained missing . the reference sequence is used as the basis for research in all clinical genomics application today.

Through the article we focus on rare disease diagnosis through sequencing. Documented rare disease phenotypes in OMIM today more than 8600. The molecular basis for 6200 of these diseases has been traced to 3900 genes in the reference genome.

Between identify the exact variants and diagnosing, we have four vital steps (sequencing, variant detection, variant assessment, and variant prioritization)

Through this article you go through steps to understand the data analysis. And also emerging machine learning and it is not an exaggeration to state that this solution could have a bearing on how the step is carried out in a diagnostic setting in the future.