Big data management challenges in healthcare

Name: Priyanshu Gupta
Roll no: 19111042
Branch: 6th Biomedical Engineering, 2022

Introduction

What is Big data?
Big Data is a term used for addressing a huge and heterogeneous amount of data too large to be handled with common software systems. The information contained by this data set can be classified into various groups and many dimensions. Such versatile databases are fragile enough and immense in volume requiring a powerful and dynamic management system including software, infrastructure to meet today's demand and privacy and security concerns. Besides these, management should be powerful enough to make it available in real time, which will allow it to be analysed and used immediately.

Big data in healthcare

A similar demand follows in the healthcare systems. In the era of telemedicine, the digitalization of medicine along with medical records and all clinical test and experiment reports is becoming a standard in hospitals. Nowadays, every single information is working through cloud and healthcare is becoming one such discipline. This information from clinical studies, research and observations of individuals' lifestyle and environmental exposure are all of great value towards advanced health research, since they work as a training datasets for predictive models for the solution and ease of real life biomedical problems. The increasing availability of omics data from a single blood test to surgeries and research reports contributing to the Big data is becoming a major challenge across health research disciplines. The main issue is the need of reliable system to deal with a dataset whose size is beyond the management capabilities of normal database software. For effective analysis knowledge extraction of correct data is one of the many challenges.

Big Data is based on three dimensions of quantity namely volume, variety, and velocity.

- The volume demands versatile storage availability and accurate approaches for retrieval of information.

- The variety allows for various combinations of grouping and different perspective for research.

- The velocity or the accelerating rate of generation of data.

Storage only represents one face of the problems. The real goal is to change the heterogeneous data into usable information and real knowledge. The Biological information is very rare and complex and so are the algorithms involved in analysing them.

Current data management practice in health research

Independently derived data archiving approaches Data management for information integration comprises two different but related practices: data organization and data preparation. The former is implemented through a database system and the latter is implemented through a data processing workflow. Both efforts will create and maintain a data archiving resource to deliver accessible and usable data products. The criteria that we used for studies include the followings: (1) the purpose and scope of a data resource; (2) data organization architecture and data preparation methods; (3) application functionality and system performance (i.e. query speed, accuracy and throughput); (4) data maintenance and governance practice and rules; (5) data readiness to use and reuse (e.g. integrity, consistency, normalization, standardization and availability of metadata).