



Big Data Usage in Medical field

Arulmani Thiyagarajan*

SRM University, School of Public Health, Chennai, India

***Corresponding Author:** Arulmani Thiyagarajan, SRM University, School of Public Health, Chennai, India.

Received: May 24, 2018; **Published:** June 14, 2018

Ongoing through the data-driven world in the present, Data is flourishing everywhere from all corners, particularly from the medical field, the quantity goes from megabytes to zettabytes. Data always remains competitive evidence in the field of health and medicine. It is predicted that the volume of data would grow 40 zettabytes by 2020 [1,2].

If remote, inaccessible and underutilized data are siloed, it would become essential and used in many ways which may have a large scope for improvement in the medical field [2]. Remarkable progress in the data storage and computing enables modern applications for new technology and services. Upcoming demand requires further advances in the process of data collection, management, and delivery - in context, in step with up driven healthcare competence and knowledge [3].

Big data as the name indicates, it has three special features in it: size or scale in terms of three v's volume, velocity, and variety, evolving, varied, distributed, timeliness, dynamic and complexity and privacy [4]. Big Data analytics discovers unknown patterns, unidentified correlations, and other understandings through probing large-scale various data sets [5]. Genomic medicine attempts to build individualized strategies for diagnostic or therapeutic decision-making by utilizing patients' genomic information [6]. While integration and manipulation of diverse genomic data and comprehensive electronic health records (EHRs) on a Big Data infrastructure exhibit challenges, they also provide a feasible opportunity to develop an efficient and effective approach to identify clinically actionable genetic variants for individualized diagnosis and therapy.

In this paper, we review what would be the role of big data in disease model prediction and drug efficacy identification, challenges of manipulating large-scale next-generation sequencing (NGS) data and diverse clinical data derived from the EHRs for genomic medicine to drug efficacy identification and disease prediction model.

There are many practical big data toolset for identifying clinically actionable genetic variants using high-throughput NGS data and HER [7].

The Software or interface used for big data analysis are Hadoop system, R Statistical language, Revolution analytics and Tableau software. Most common Database used in big data analysis is DBMS, RDBMS, and ORDBMS. Big data science may overcome some challenges in the evidence-based medical system as this system used large data which is of multidimensional in nature, through which numerous evidence-based information can be identified and used in medical research purposes [8]. It allows predictive analytics to understand not only what has happened and what is currently happening, but also to predict what will happen in the future [9].

The main key challenges in the big data analysis are that software with lack of adaptive intelligent tools, accessibility and appropriate training at the current stages [10].

The basic approaches for any big data analysis are of following (Questions to ask before approaching)

1. How to turn big data into good research problems/questions/hypothesis, then transform into valuable solutions that benefit us and follow basic steps (Figure 1)
2. How to convert zeta byte data to quantitative statistical evidence for diagnostics, therapeutics, and new acumens into population health, disease and treatment [11].
3. What are the best approaches?
4. Do traditionally used interference techniques continue to play some roles?
5. Experimental versus computational?
6. Hypothesis or data driven?
7. Traditional statistical modeling or data mining and artificial approaches?
8. Batch processing or individual clustering?

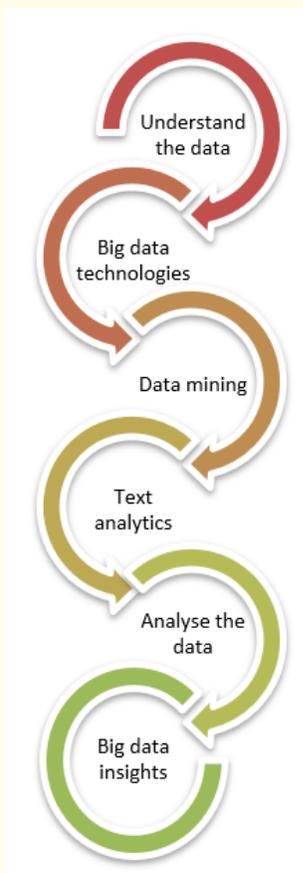


Figure 1: Steps of Big data analysis.

Analysis by Bottleneck procedure may scale up with the big data with its complexity by involving advanced statistical and computational techniques with pipelines [12]. Validation, interpretation, and visualization are the most important concepts that need to be done for any prediction or drug effectiveness or intervention effect. One health paradigm is a concept in which disease occurrence (biological event) in humans, animals, and plants.

Disease prediction modeling helps us in the following fields: Bio-Surveillance, Remote Sensing, Risk assessment, spatial epidemiology and Ecological niche. Models were classified as either validated and verified (V&V) or not V&V [13]. The method used for validation and verification were: Statistical verification, sensitivity analysis, specificity and sensitivity, verification using training data, validation using temporally independent data and validation using spatially and temporal independent data [13]. Most frequently used verification is Statistical verification involving kappa statistics, Receiver Operating Characteristic curve- Goodness of fit.

Biomarkers or novel clinical biochemical parameter is used for assessing the disease prediction. Clinical prediction help with associated decision making which facilitates patient-doctor communication based on the more objective information. A clinical prediction model is used to explore the relationship between future or unknown outcomes and baseline health states among people with specific conditions [14]. Clinical prediction model mostly based on the evidence-based medicine, but researcher must follow the recommended guidelines and approaches [15]. A model should Predict specific events accurately, be relatively simple and easy to use.

Initially, Framingham study which attempted to develop tools for prediction using Discriminant function analysis for Multivariate coronary heart disease risk model, later more models have emerged from different authors including PROCAM, ASSIGN, SCORE and QRISK [16]. Diabetes prediction models have also been into existence for many years, started from San Antonio Study using multiple logistic regression

Three types of prediction models have been described [diabetes prediction model],

1. By non-invasive measures only
2. By biochemical measures, novel biomarkers
3. By genetic information

Two general model: Diagnostic prediction model, and Prognostic prediction model. Predictor – clearly defined, measurable in a standardized and reproducible way (can affect the precision of the estimates of the predictor-outcome associations). Outcome – Measurable in a standardized and consistent way across settings [1,16,17]. Missing data managed by multiple data imputation. In the context of transportable (model must provide an accurate estimate of the risk in a different but plausibly related population which is external validity. No strict indication of this validity to be performed. Types of validity are temporal validation, Geographical validation, Domain validation, Spectrum validation, and Follow-up period validation.

Instead of developing a new prediction model, the strong emphasis should be given on validation and implementation of existing prediction model which is suggested in AP Kengne., *et al*.

In case of genetic disease intervention or identification of drug effectiveness, the key biological questions would be of

1. Identification of significantly differentially expressed genes responding to the treatment

2. Characterizing the dynamics and changes of gene expression to determine the trajectories of significantly regulated genes in responding to the treatment [4].

The world is transforming with its peak of data flourishing from all different fields which in turn makes it to be in the state of data-driven. Advances in the density of computing power and data storage enable us to have better solution and platform to use big data well in the Medical field and disease model prediction. The demand for big data in context to business workflows and in the stream of life will make us develop innovative ideas to drive further in process of collection, management, and delivery of data.

Conflict of interest

There is no conflict of Interest.

Bibliography

1. "Big data Analytics and Predictive Analytics - Editor Review, User Reviews, Features, Pricing and Comparison in 2018". *Predictive Analytics Today* (2015).
2. J Gantz and D Reinsel. "The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East - United States" (2013): 7.
3. E Check Hayden. "Genome researchers raise alarm over big data". *Nature News* (2015).
4. Y Liang and A Kelemen. "Big Data Science and Its Applications in Health and Medical Research: Challenges and Opportunities". *Journal of Biometrics and Biostatistics* 7.3 (2016): 307.
5. KY He., et al. "Big Data Analytics for Genomic Medicine". *International Journal of Molecular Sciences* 18.2 (2017).
6. K He., et al. "Big Data Analytics for Genomic Medicine". *International Journal of Molecular Sciences* 18.2 (2017): E412.
7. N Elgendy and A Elragal. "Big Data Analytics: A Literature Review Paper". In *Advances in Data Mining. Applications and Theoretical Aspects* (2014): 214-227.
8. AS Tiwarkhede and V Kakde. "A Review Paper on Big Data Analytics". *International Journal of Science and Research* 4.4 (2013): 4.
9. HS Bhosale and DP Gadekar. "A Review Paper on Big Data and Hadoop". *International Journal of Scientific and Research Publications* 4.10 (2014): 7.
10. J Fan., et al. "Challenges of Big Data analysis". *National Science Review* 1.2 (2014): 293-314.
11. V Kargin. "Statistical properties of zeta functions' zeros". *Probability Surveys* 11.0 (2014): 121-160.
12. CL Philip Chen and CY Zhang. "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data". *Information Sciences* 275 (2014): 314-347.
13. CD Corley., et al. "Disease Prediction Models and Operational Readiness". *PLOS ONE* 9.3 (2014): e91989.
14. Y Lee., et al. "How to Establish Clinical Prediction Models". *Endocrinology and Metabolism* 31.1 (2016): 38-44.
15. DL Sackett., et al. "Evidence based medicine: what it is and what it isn't". *British Medical Journal* 312.7023 (1996): 71-72.
16. S van Dieren., et al. "Prediction models for the risk of cardiovascular disease in patients with type 2 diabetes: a systematic review". *Heart* 98.5 (2012): 360-369.
17. JB Echouffo-Tcheugui., et al. "Risk Models to Predict Hypertension: A Systematic Review". *PLoS ONE* 8.7 (2013): e67370.

Volume 2 Issue 4 July 2018

© All rights are reserved by Arulmani Thiyagarajan.