Improving Healthcare Using Big Data

Mohana Priya.A[1], Kavya.S.P[2], Aruna devi.S[3], Dhivya shri.P[4], Logeswari.K[5]
[1, 2] Assistant Professor, Dept of Computer Science And Engineering
[1, 2, 3, 4, 5] Sri Shakthi Institute Of Engineering And Technology, Coimbatore

Abstract- *The fastly growing field of big data analytics has started to play a important role in the evolution of healthcare. It has provided tools to manage, analyze, structured, and unstructured data produced by current healthcare systems. In this paper, we are going to discuss the major challenges that include three promising areas of medical research: image, signal, and genomics based analytics. Potential areas of research within this field which have the ability to provide meaningful impact on healthcare.*

Keywords- Big data; Analytics; Healthcare; Analytical tools.

I. INTRODUCTION

Big Data plays a vital role in the field of healthcare. Big Data have characteristics like volume, velocity, variety and veracity. Since the information is increasing vastly day-by-day, Big Data not only defines the size but also finds insights from unstructured, complex, noisy, heterogeneous, longitudinal and voluminous data. It aims to aggregate years of medical data into medical databases. Big Data analytics can bring the revolution in healthcare industry. This data in healthcare provide opportunity to perform predictive analysis. It has a great potential to process large amount of data in parallel and it can also used to find the solution of hidden problems . For example, any disease that has occurred earlier in any parts of the world, pre-diction of that disease can be done efficiently. Although, clinics and hospitals can reprocess the data to analyze and calculate the patient preferences. In predictive analysis, we implement different statistical methods, data mining, and machine learning approaches to analyze, process, and predict the conclusion for undiscovered data. The lot of possibilities to provide better cure for disease using different analytical tools in healthcare domain. Big data may include structured, semi-structured, and unstructured data. First, we collect data that is generated from different origins and then collect and store it into one common platform.

II. THREE AREAS OF BIG DATA ANALYTICS IN MEDICINE

- Image Processing
- Signal Processing
- Genomics

Image Processing.

Medical images are an important source of data periodically used for diagnosis, therapy assessment and planning. Computed tomography (CT), magnetic resonance imaging (MRI), X-ray, molecular imaging, ultra-sound, photo acoustic imaging, fluoroscopy, positron emission tomography-computed tomography (PET-CT), and mammography are some of the examples of imaging techniques that are well established within clinical settings. Medical image data can range from a few megabytes for a single study (e.g., histology images) to hundreds of megabytes per study. Such data requires large storage capacities if stored for long term. It also demands fast and accurate algorithms if any decision assisting automation were to be performed using the data. If other sources of data acquired for each patient are applied during the diagnoses, prognosis, and treatment processes. Then the problem of providing cohesive storage and developing efficient methods capable of encapsulating the wide range of data becomes a challenge.

Signal Processing

Medical signals are used especially during high-resolution acquisition, continuous and storage from a multitude of monitors. These are connected to each patient and also pose volume and velocity obstacles. It is similar to medical images. In addition to the data size issues, physiological signals also pose complication of a spatiotemporal nature. Analysis of physiological signals, when presented along with situational context awareness is often more meaningful. This context needs to be embedded into the development of continuous monitoring and predictive systems to ensure its effectiveness and robustness.

The current healthcare systems use numerous disparate and continuous monitoring devices. These devices utilize singular physiological waveform data or discretized vital information to provide alert mechanisms in case of overt events. The alarm systems are developed and implemented that are tend to be reliable in simple ways. Their sheer numbers could cause alarm fatigue for both care givers and patients. The new medical knowledge is determined from the ability of this setting. It is constrained by prior knowledge that has typically fallen short of maximally utilizing high-dimensional time series data. The reasons behind the failure of alarm mechanisms are these systems rely on single sources of information while lacking context of the patients' true physiological conditions from a broader and more comprehensive viewpoint. So, we need to develop more improved and comprehensive approaches based on the study of interactions and correlations among multimodal clinical time series data. It is important to show that the humans are poor in reasoning about changes affecting more than two signals.

Genomics

The cost to sequence the human genome is rapidly decreasing with the development of high-throughput sequencing technology. With implications for current public health policies and delivery of care, analyzing genome-scale data for developing actionable recommendations in a timely manner is a significant challenge to the field of computational biology. The crucial part is cost and time to deliver the recommendations in a clinical setting. Initiatives tackling this complex problem include tracking of 100,000 subjects over 20 to 30 years using the predictive, preventive, participatory, and personalized health, referred to as P4, medicine paradigm as well as an integrative personal omics profile. The P4 initiative is using a system approach for

i. to determine disease states by analyzing genome-scale datasets.
ii. moving towards blood based diagnostic tools for continuous monitoring of a subject.
iii. exploring new approaches to drug target discovery, developing tools to deal with big data challenges of capturing, validating, storing, mining, integrating, and finally
iv. modeling data for each individual. The integrative personal omics profile combines physiological monitoring and multiple high-throughput methods for genome sequencing to generate a detailed health and disease states of a subject. Ultimately, realizing actionable recommendations at the clinical level remains a grand challenge for this field. Utilizing such high density data for exploration, discovery, and clinical translation demands novel big data approaches and analytics.

v. Despite the enormous expenditure consumed by the current healthcare systems, clinical outcomes remain suboptimal, particularly in the USA, where 96 people per 100,000 die annually from conditions considered treatable.

III. BIG DATA APPLICATIONS IN GENOMICS

The advent of high-throughput sequencing methods has enabled researchers to study genetic markers over a wide range of population, improve efficiency by more than five orders of magnitude since sequencing of the human genome was completed, and associate genetic causes of the phenotype in disease states. Genome-wide analysis utilizing microarrays has been successful in analyzing traits across a population and contributed successfully in treatments of complex diseases such as Crohn's disease and age-related muscular degeneration.

Analytics of high-throughput sequencing techniques in genomics is an inherently big data problem as the human genome consists of 30,000 to 35,000 genes. Initiatives are currently being pursued over the timescale of years to integrate clinical data from the genomic level to the physiological level of a human being. These initiatives will help in delivering personalized care to each patient. Delivering recommendations in a clinical setting requires fast analysis of genome-scale big data in a reliable manner. This field is still in a nascent stage with applications in specific focus areas, such as cancer, because of cost, time, and labour intensive nature of analyzing this big data problem.

Big data applications in genomics cover a wide variety of topics. Here we focus on pathway analysis, in which functional effects of genes differentially expressed in an experiment or gene set of particular interest are analysed, and the reconstruction of networks, where the signals measured using high-throughput techniques are analysed to reconstruct underlying regulatory networks. These networks influence numerous cellular processes which affect the physiological state of a human being.

IV. PREDICTIVE ANALYTICS USING BIG DATA

In the accelerating, the adoption of the technique states that the volume and detail of patient information is growing rapidly. The surge in the creation and broadening use of technique was driven in part by a $30 billion federal government stimulus, provided by the organisation. The organisation was designed specifically to provide incentives to adopt the technique and this organisation encourage the sharing of patient details by clinicians everywhere in an attempt to lower costs, speed diagnosis, and improve patient outcomes. The combination of different types of structured and unstructured data are also analysed across multiple data sources. Accurate rate of diagnosing patient conditions, matching treatments with outcomes and predicting patients at risk for disease. Predictive modelling over data derived from techniques is being used for early diagnosis and is reducing mortality rates from problems such as congestive heart failure and sepsis. Healthcare spent more in Congestive Heart Failures. By avoiding expensive complications, physicians will be easily missed by previous demonstration. By adding additional features AUC(area under curve), there was a substantial increase in the ability of the model to distinguish people who have Congestive Heart Failure from people and machine learning example from an organisation demonstrated that machine learning algorithms could look at many more factors in patient's charts than doctors. (Fig: 1). .

More Data points Improve Heart Disease Diagnosis

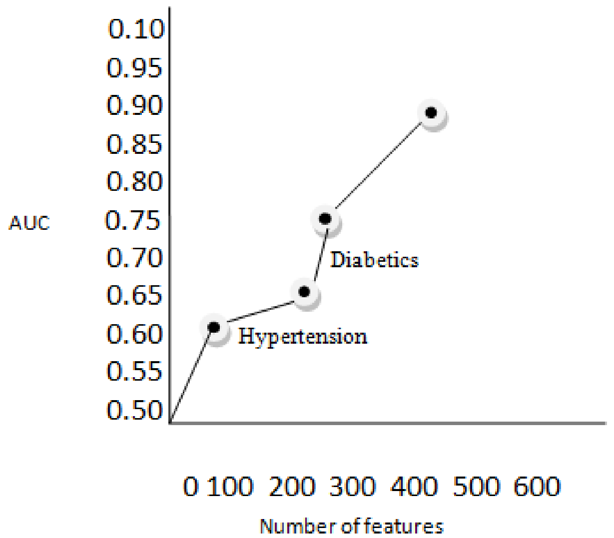

Fig: 1

V. CONCLUSION

The future evolution in healthcare plays a vital role in big data analytics which contains disparate, structured, and unstructured data sources. One can already see a spectrum of analytics being used, aiding in the decision making and performance of healthcare personnel and patients. We mainly focused on three areas: medical image analysis, physiological signal processing, and genomic data processing. The scientists come up with innovative solutions to process this large volume of data in manageable timescales in the growth of the volume of medical images. The development of imaginative and incredible systems that help save lives for physiological signal processing from both research and practicing medical professionals that are growing steadily. In the development of a detailed model of a human being, the combination of physiological data and high-throughput "omics" techniques were used. That has the potential to enhance our knowledge of diseases and help in the development of blood related diagnostic tools. The similar challenges and opportunities in dealing with disparate, structured and unstructured big data sources are: medical image analysis, signal processing of physiological data, and integration of physiological and "omics" data.

REFERENCES

[1] F. Ritter, T. Boskamp, A. Homeyer et al., "Medical image analysis," *IEEE Pulse*, vol. 2, no. 6, pp. 60–70,2011.

[2] M. Cvach, "Monitor alarm fatigue: an integrative review," *Biomedical Instrumentation & Technology*, vol. 46, no. 4, pp. 268–277, 2012.

[3] R. Drmanac, A. B. Sparks, M. J. Callow et al., "Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays," *Science*, vol. 327, no. 5961, pp. 78–81, 2010.

[4] R. Chen, G. I. Mias, J. Li-Pook-T han et al., "Personal omics pro-filing reveals dynamic molecular and medical phenotypes," *Cell*, vol. 148, no. 6, pp. 1293–1307, 2012.

[5] D. Freeman, "The future of patient monitoring," *Health Management Technology*, vol. 30, no. 12, article 26, 2009.