International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018IJSRCSEIT| Volume 4 | Issue 3 | ISSN: 2456-3307

Big Data Analytics in Healthcare

Saranya K[1], UdhayaKumar U[2]
[1]Department of Computer Science, Shanmuga Industries Arts and Science College, Tiruvannamalai, Tamilnadu, India
[2]Assistant Professor, Department of Computer Science, Shanmuga Industries Arts and Science College, Tiruvannamalai, Tamilnadu, India

ABSTRACT

The rapidly expanding field of big data analytics has started to play a pivotal role in the evolution of healthcare practices and research. It has provided tools to accumulate, manage, analyse, and assimilate large volumes of disparate, structured, and unstructured data produced by current healthcare systems. Big data analytics recently applied towards aiding the process of care delivery and disease exploration. In this paper, we discuss some of these significant challenges with a focus on three upcoming and promising areas of medical research: image, signal, and genomics-based analytics.

Keywords: Big Data, Big Data Analytics, 3V's, Image Processing, Signal processing, Genomics.

I. INTRODUCTION

The concept of "big data" is not new; however, the way it is defined is continually changing. Various attempts at defining big data essentially characterize it as a collection of data elements whose size, speed, type, and/or complexity require one to seek, significant, and invent new hardware and software mechanisms in order to successfully store, analyse, and visualize the data. Healthcare is a prime example of how the three V's of data, velocity (speed of generation of data), variety, and volume is an innate aspect of the data it produces.

Historical approaches to medical research have focused on the investigation of disease states based on the changes in physiology in the form of a confined view of the specific singular modality of data Important physiological and pathophysiological phenomena are concurrently manifest as changes across multiple clinical streams. This results from strong coupling among different systems within the body (e.g., interactions between heart rate, respiration, and blood pressure) thereby producing potential markers for clinical assessment.

In this paper, three areas of big data analytics in medicine discussed. These three areas do not comprehensively reflect the application of big data analytics in medicine.

Image Processing: Medical images are an essential source of data frequently used for diagnosis, therapy assessment, and planning. Computed tomography (CT), magnetic resonance imaging (MRI), X-ray, molecular imaging, ultrasound, are some of the examples of imaging techniques.

Signal Processing: Similar to medical images, medical signals also pose volume and velocity obstacles especially during continuous, high-resolution acquisition and storage from a multitude of monitors connected to each patient. Currently, healthcare systems use numerous different and continuous monitoring devices that singular physiological waveform data or discretised vital information to provide alert mechanisms in case of open events.

Genomics: The cost to sequence the human genome (encompassing 30,000 to 35,000 genes) is rapidly decreasing with the development of high-throughput sequencing technologythe 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) analysing genome-scale datasets to determine disease states, (ii) moving towards blood-based diagnostic tools for continuous monitoring of a subject, (iii) exploring new approaches to drug targetdiscovery, developing tools to deal with big data challenges of capturing, validating, storing, mining, integrating, and finally (iv) modelling data for each.

II. MEDICAL IMAGE PROCESSING FROM BIG DATA POINT OF VIEW

Medical imaging provides essential information on anatomy and organ function in addition to detecting diseases states. Moreover, it is utilized for organ delineation, identifying

Tumors in lungs, spinal deformity diagnosis, artery stenosis detection, aneurysm detection, and so forth. In these applications, image processing techniques such as enhancement, segmentation, and denoising in addition to machine learning methods employed. As the size and dimensionality of data increase, understanding the dependencies among the data and designing efficient, accurate, and computationally efficient methods demand new computer-aided techniques and platforms. The rapid growth in the number of healthcare organizations as well as the number of patients has resulted in the excellent use of computer-aided medical diagnostics and decision support systems in clinical settings.

Many areas of health care such as diagnosis, prognosis, and screening can improve by utilizing computational intelligence. The integration of computer analysis with appropriate care has potential to help clinicians improve diagnostic accuracy. The integration of medical images with other typesofelectronic health record (EHR) data and genomic data can also improve the accuracy and reduce the time taken for diagnosis. In the following, data produced by imaging techniques reviewed and applications of medical imaging from a big data point of view are discussed.

A. Data Produced by Imaging Techniques

Medical imaging encompasses a broad spectrum of different image acquisition methodologies typically utilized for a variety of clinical applications. For example, visualizing blood vessel structure can be performed using magnetic resonance imaging (MRI), computed tomography (CT), ultrasound, and photoacoustic imaging [30]. From a data dimension point of view, medical images might have 2, 3, and four dimensions. Positron emission tomography (PET), CT, 3D ultrasound, and functional MRI (fMRI) considered as multidimensional medical data. Modern medical image technologies can produce higher solution images such as respiration-correlated or “four-dimensional” computed tomography (4D CT) Molecular imaging is a noninvasive technique of cellular and subcellular events which has the potential for clinical diagnosis of disease states such as cancer. However, to make it clinically applicable for patients, the interaction of radiology, nuclear medicine, and biology is crucial that could complicate its automated analysis. Microwave imaging is an emerging methodology that could create a map of electromagnetic wave scattering arising from the contrast in the dielectric properties of different tissues. It has both functional and physiological information encoded in the dielectric properties which can help differentiate and characterize different tissues and pathologies. However, microwaves have scattering behavior that makes retrieval of data a challenging task. Advanced

Multimodal Image-Guided Operating (AMIGO) suite designed which has angiographic X-ray system, MRI, 3D ultrasound, and PET/CT imaging in the operating room (OR). This system has been used for cancer therapy and showed the improvement in localization and targeting an individual's diseased tissue [1, 2].

B. Methods

The volume of medical images is growing exponentially. For instance, Image CLEF medical image dataset contained around 66,000 images between 2005 and 2007.

While just in the year of 2013 around 300,000 images were stored every day [3]. In addition to the growing volume of images, they differ in modality, resolution, dimension, and quality, which introduce new challenges such as data integration and mining especially if multiple datasets are involved. Compared to the volume of research that exists on single modal medical image analysis, there is the considerably lesser number of research initiatives on multimodal image analysis.

C. Analytical Methods

The goal of medical image analytics is to improve the interpretability of depicted contents [4].Many methods and frameworks developed for medical image processing. However, these methods are not necessarily applicable for big data applications. One of the frameworks developed for analyzing and transformation of huge datasets is Hadoop that employs Map Reduce Map Reduce is a programming paradigm that provides scalability across many servers in a Hadoop cluster with a broad variety of real-world applications. However, it does not perform well with input output intensive tasks Map Reduce framework.It has been used in [5] to increase the speed of three large-scale medical image processing use-cases, (i) finding optimal parameters for lung texture classification by employing a well-known machine learning method, support vector machines (SVM), (ii) content-based medical image indexing, and (iii) wavelet analysis for robust texture classification.

D. Collecting, Sharing, and Compressing Methods

In addition to developing analytical methods, efforts have made for collecting, compressing, sharing, and anonymizing medical data. One example is IDASH (integrating data for analysis, anonymization, and sharing) which is a center for biomedical computing [6]. It focuses on algorithms and tools for sharing data in a privacy-preserving manner. The goal of iDASH is to bring together a multi-institutional team of quantitative scientists to develop algorithms and tools, services, and a biomedical cyberinfrastructure to be used by biomedical and behavioral researchers [7]. Another example of a similar approach is Health-e-Child consortium of 14 academic, industry, and clinical partners with the aim of developing an integrated healthcare platform for European pediatrics [8].

There are some limitations in implementing the application-specific compression methods on both general purpose processors and parallel processors such as graphics processing units (GPUs) as these algorithms need highly variable control and the complex bit well suited to GPUs and pipeline architectures. To overcome this limitation, an FPGA implementation proposed for LZ-factorization which decreases the computational burden of the compression algorithm [9]. Lossy image compression has been introduced in [10] that reshapes the image in such a way that if the image is uniformly sampled, sharp features have a higher sampling density than the rough ones. This method is claimed to be applicable for big data compression. However, for medical applications lossy methods are not applicable in most cases as fidelity is essential and information must preserve. These techniques are among a few techniques that have been either designed as prototypes or developed with limited applications. Developing methods for processing/analysing a broad range and large volume of data with acceptable accuracy and speed is still critical. In Table 1, we summarize the challenges facing medical image processing. When dealing with big data, these challenges seemed to be more severe and on the other hand analytical methods could benefit the big data to handle them.

III. MEDICAL SIGNAL ANALYTICS

Streaming data analytics in healthcare can be defined as a systematic use of continuous waveform (the signal varying against time) and related medical record information developedthrough applied analytical disciplines (e.g., statistical, quantitative, contextual, cognitive, and predictive) to drive decision making for patient care. The analytics workflow of real-time streaming waveforms in clinical settings can be broadly described.

Firstly, a platform for streaming data acquisition and ingestion is required, which has the bandwidth to handle multiple waveforms at different fidelities. Integrating these dynamic waveform data with static data from the EHR is a crucial component to provide situational and contextual awareness for the analytics engine. Enriching the data consumed by analytics not only makesthe system more robust but also helps balance the sensitivity and specificity of the predictive analytics. The specifics of the signal processing will largely depend on the type of disease cohort under investigation. A variety of signal processing mechanisms can be utilized to extract a multitude of target features which are then consumed by a trained machine learning model to produce actionable insight.

These actionable insights could either be diagnostic, predictive, or prescriptive. These insights could further be designed to trigger other mechanisms such as alarms and notification to physicians.

A. Data Acquisition

Historically streaming data from connected physiological signal acquisition devices rarely stored. Even if the option to store this data were available, thelength of these data captures was typically short and downloaded only using proprietary software and data formats provided by the device manufacturers. Although most major medical device manufacturers are now taking steps to provide interfaces to access live streaming data from their devices, such data in motion very quickly poses archetypal big datachallenges. The fact that there are also governance challenges such as lack of data protocols, lack of data standards, and data privacy issues are adding to this.

B. Data Storage and Retrieval

With massive volumes of streaming data and other patient information that can gather from clinical settings, sophisticated storage mechanisms of such data are imperative. Since storing and retrieving can be computational and time expensive, it is critical to have a storage infrastructure that facilitates rapid data pull and commits based on analytic demands. With its capability to store and compute large volumes of data, usage of systems such as Hadoop, Map Reduce, and Mongo DB [11, 12] is becoming much more familiar with the healthcare research communities. Mongo DB is a free cross-platform document-oriented database

C. Data Aggregation

Integration of disparate sources of data, developing consistency within the data, standardization of data from similar sources, and improving the confidencein the data especially towards utilizing automated analytics are among challenges facing data aggregation in healthcare systems. Analysis of continuous data heavily utilizes the information in time domain. However, static data does not always provide accurate time context and, hence, when combining the waveform data with static electronic health record data, thetemporal nature of the time context during integration can also add significantly to the challenges. There are considerable efforts in compiling waveforms and other associated electronic medical information into one cohesive database that are made publicly available to researchers worldwide [13]. For example, MIMIC II [14] and some other datasets included in Physionet [15] provide waveforms and other clinical data from a wide variety of actual patient cohorts.

D. Signal Analytics Using Big Data

Research in signal processing for developing big data-based clinical decision support systems (CDSSs) is getting more prevalent. In fact, organizations such as the Institution of Medicine have long advocated use of health information technology including CDSS to improve care quality CDSSs provide medical practitioners with knowledge and patient-specific information, intelligently filtered and presented at appropriate times, to improve the delivery of care. A study presented by Lee and Mark uses the MIMIC II database to prompt therapeutic intervention to hypotensive episodes using cardiac and blood pressure time series data another study shows the use of physiological waveform data along with clinical data from the MIMIC II database for finding similarities among patients within the selected cohorts.

As complex physiological monitoring devices are getting smaller, cheaper, and more portable, personal monitoring devices are being used outside of clinical environments by both patients and enthusiasts alike. However, similar to clinical applications, combining information simultaneously collected from multiple portable devices can become challenging.Pantelopoulos and Bourbakis discussed the research and development of wearable biosensor systems and identified the advantages and shortcomings in this area of study. Similarly, portable and connected electrocardiogram, blood pressure and body weight devices used to set up a network-based study of telemedicine. The variety of fixed as well as mobile sensors available for data mining in the healthcare sector and how such data can be leveraged for developing patient care technologies are surveyed.

IV. 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 to 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.

A. Pathway Analysis

Resources for inferring functional effects for "-omics" big data are largely based on statistical associations between observed gene expression changes andpredicted functional effects. Experiment and analytical practices lead to error as well as batch effects Interpretation of functional effects has to incorporate continuous increases in available genomic data and corresponding annotation of genes.There are variety of tools, but no "gold standard" for functional pathway analysis of high-throughput genome-scale data. Three generations of methods used for pathway analysis are described as follows. The first generation encompasses overrepresentation analysis approaches that determine the fraction of genes in a particular pathway found among the genes, which are differentially expressed. Examples of the first generation tools are Onto-Express Go Miner and Clue Go. The second generation includes functional class scoring approaches, which incorporate expression levelchanges in individual genes as well as functionally similar genes.

B. Reconstruction of Regulatory Networks

Pathway analysis approaches do not attempt to make sense of high-throughput big data in biology as arising from the integrated operation of a dynamical system. There are multiple approaches to analyzing genome-scale data using a dynamical system framework [30, 31]. Due to the breadth of the field, in this section, we mainly focus on techniques to infer network models from big biological data. Reconstruction of metabolic networks has advanced in last two decades. One objective is to develop an understanding of organism-specific metabolism through reconstruction of metabolic networks by integrating genomics, transcriptomics, and proteomics high-throughput sequencing techniques [33]. Constraint-based methods are widely applied to probe the genotype-phenotype relationship and attempt to overcome the limited availability of kinetic constants Reconstruction of gene regulatory networks from gene expression data is another well-developed field.

Network inference methods can be split into five categories based on the underlying model in each case: regression, mutual information, correlation, Boolean regulatory networks, and other techniques [16, 35]. Over 30 inference techniques were assessed after DREAM5 challenge in 2010 [17, 34]. Boolean regulatory networks [18] are a particular case of discrete dynamical models where the state of a node or a set of nodes exists in a binary state. The actual state of each node or set of nodes is determined by using Boolean operations on the state of other nodes in the network [19]. Boolean networks are beneficial when an amount of quantitative data is small [10, 20] but yield the high number of false positives (i.e., when a given condition is satisfied while actually, that is not the case) that may be reduced by using prior knowledge [21, 22]. Another bottleneck is that Boolean networks are prohibitively expensive when the number of nodes in the network is large. This is due to the number of global states rising exponentially in the number of entities [23].A method to overcome this bottleneck is to use clustering to break down the problem size. For example, Martin et al. [24] broke down a 34,000-probe microarray gene expression dataset into 23 sets of metagenes using clustering techniques. This Boolean model successfully captured the network dynamics for two different immunology microarray datasets. The dynamics of gene regulatory network can be captured using ordinary differential equations (ODEs) [12-15]. This approach has been applied to determine regulatory network for yeast [25]. The study successfully captured the regulatory network which has been characterized using experiments by molecular biologists. Reconstruction of a gene regulatory network on a genome-scale system as a dynamical model is computationally intensive [26]. A parallelizable dynamical ODE model has been developed to address this bottleneck [27]. It reduces the computational time to O(N2) from the time taken in other approaches which are O(N3) or O(N2 log N) [28].Determining connections in the regulatory network for a problem of the size of the human genome, consisting of 30,000 to 35,000 genes [16, 17], will require exploring close to a billion possible connections. The dynamical ODE model has been applied to reconstruct the cardiogenic gene regulatory network of the mammalian heart [29].

V. CONCLUSION

Big data analytics which leverages legions of disparate, structured and unstructured data sources is going to play a vital role in how healthcare practiced in the future. One can already see a spectrum of analytics being utilized, aiding in the decision-making and performance of healthcare personnel and patients. Here we focused on three areas of interest: medical image analysis, physiological signal processing, and genomic data processing.

The acquisition, formation/reconstruction, enhancement, transmission, and compression. New technological advances have although there are some genuine challenges for signal processing of physiological data to deal with, given the current state of data competency and no standardized structure, there are opportunities in each step of the process towards providing systemic improvements within the healthcare research and practice communities. Apart from the apparent need for further research in the area of data wrangling, aggregating, and harmonizing continuous and discrete medical data formats, there is also an equal need for developing novel signal processing techniques specialized towards physiological signals. Research about mining for biomarkers and hidden patterns within bio signals tounderstand and predict disease cases has shown potential in providing actionable information. However, there are opportunities for developing algorithms to address data filtering, interpolation, transformation, feature extraction, feature selection, and so forth. Furthermore, with the notoriety and improvement of machine learning algorithms, there are opportunities in improving and developing robust CDSS for clinical prediction, prescription, and diagnostics.

Integration of physiological data and high-throughput "- omics" techniques to deliver clinical recommendations is the grand challenge for systems biologists. Although associatingFunctional effects with changes in gene expression have progressed, the continuous increase in available genomic data and its corresponding impact of annotation of genes and errors from experiment and analytical practices make analyzing functional effect from high-throughput sequencing techniques a challenging task. Reconstruction of networks on the genome-scale is an ill-posed problem. Robust applications developed for reconstruction of metabolic networks and gene regulatory networks. Limited availability of kinetic constants is a bottleneck, and hence various models attempt to overcome this limitation. There is an incomplete understanding of this large-scale problem as gene regulation, an effect of different network architectures, and evolutionary effects on these networks are still being analysed [135]. To address these concerns, the combination of the careful design of experiments and model development for reconstruction of networks will help in saving time and resources spent in building the understanding of regulation in genome-scale networks. The opportunity of addressing the grand challenge requires close cooperation among experimentalists, computational scientists, and clinicians.

VI.REFERENCES

[1] A. McAfee, E. Brynjolfsson, T. H. Davenport, D. J. Patil, and D. Barton, "Big data: the management revolution," *Harvard Business Review*, vol. 90, no. 10, pp. 60–68, 2012.

[2] C. Lynch, "Big data: how do your data grow?" *Nature*, vol. 455,no. 7209, pp. 28–29, 2008.

[3] A. Jacobs, "The pathologies of big data," *Communications of the ACM*, vol. 52, no. 8, pp.

[4] P. Zikopoulos, C. Eaton, D. deRoos, T. Deutsch, and G. Lapis, *Understanding Big Data: Analytics for Enterprise Class Hadoopand Streaming Data*, McGraw-Hill Osborne Media.

[5] J.Manyika,M. Chui, B. Brown et al., Big Data:The Next Frontierfor Innovation, Competition, and Productivity,McKinsey Global Institute, 2011.

[6] J. J. Borckardt, M. R. Nash, M. D. Murphy, M. Moore, D. Shaw, and P. O'Neil, "Clinical practice as natural laboratory for psychotherapy research: a guide to case-based time-series analysis," *The American Psychologist*, vol. 63, no. 2, pp. 77–95, 2008.

[7] L. A. Celi, R.G.Mark, D. J. Stone, and R.A.Montgomery, "'Big data' in the intensive care unit: closing the data loop,"

AmericanJournal of Respiratory andCriticalCareMedicine, vol. 187, no. 11,pp. 1157–1160, 2013.

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

[9] J. A. Seibert, "Modalities and data acquisition," in *PracticalImaging Informatics*, pp. 49–66, Springer, New York, NY, USA, 2010.

[10] B. J. Drew, P. Harris, J. K. Z`egre-Hemsey et al., "Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients," *PLoSONE*, vol. 9, no. 10, Article IDe110274, 2014. BioMed Research International

[11] K. C. GrahamandM. Cvach, "Monitor alarmfatigue: standardizing use of physiological monitors and decreasing nuisance alarms,"*The American Journal of Critical Care*, vol. 19, no. 1, pp. 28–34, 2010.

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

[13] J. M. Rothschild, C. P. Landrigan, J. W. Cronin et al., "The Critical Care Safety Study: the incidence and nature of adverse events and serious medical errors in intensive care," *CriticalCare Medicine*, vol. 33, no. 8, pp. 1694–1700, 2005.

[14] P. Carayon and A. P. G¨urses, "A human factors engineering conceptual framework of nursing workload and patient safety in intensive care units," *Intensive and Critical Care Nursing*, vol.21, no. 5, pp. 284–301, 2005.

[15] P.Carayon, "Human factors of complex sociotechnical systems," *Applied Ergonomics*, vol. 37, no. 4, pp. 525–535, 2006.

[16] E. S. Lander, L. M. Linton, B. Birren et al., "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001.

[17] 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.

[18] T. Caulfield, J. Evans, A.McGuire et al., "Reflections on the cost of 'Low-Cost' whole genome sequencing: framing the health policy debate," *PLoS Biology*, vol. 11, no. 11, Article ID e1001699,2013.

[19] F. E. Dewey, M. E. Grove, C. Pan et al., "Clinical interpretation and implications of whole-genome sequencing," *JAMA*, vol. 311, no. 10, pp. 1035–1045, 2014.

[20] L. Hood and S. H. Friend, "Predictive, personalized, preventive, participatory (P4) cancer medicine," *Nature Reviews ClinicalOncology*, vol. 8, no. 3, pp. 184–187, 2011.

[21] L. Hood and M. Flores, "A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory," *New Biotechnology*, vol. 29, no. 6, pp. 613–624, 2012.

[22] L. Hood and N. D. Price, "Demystifying disease, democratizing health care," *Science Translational Medicine*, vol. 6, no. 225, Article ID 225ed5, 2014.

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

[24] G. H. Fernald, E. Capriotti, R. Daneshjou, K. J. Karczewski, and R. B. Altman, "Bioinformatics challenges for personalized medicine," *Bioinformatics*, vol. 27, no. 13, Article ID btr295, pp.1741–1748, 2011.

[25] P. Khatri, M. Sirota, and A. J. Butte, "Ten years of pathway analysis: current approaches and outstanding challenges," *PLoSComputational Biology*, vol. 8, no. 2, Article ID e1002375, 2012.

[26] J. Oyelade, J. Soyemi, I. Isewon, and O. Obembe, "Bioinformatics, healthcare informatics and analytics: an imperative for improved healthcare system," *International Journal of*

AppliedInformation Systems, vol. 8, no. 5, pp. 1–6, 2015.

[27] T. G. Kannampallil, A. Franklin, T. Cohen, and T. G. Buchman, “Sub-optimal patterns of information use: a rational analysis of information seeking behavior in critical care,” in *CognitiveInformatics in Health and Biomedicine*, pp. 389–408, Springer, London, UK, 2014.

[28] H. Elshazly, A. T. Azar, A. El-korany, and A. E. Hassanien, “Hybrid system for lymphatic diseases diagnosis,” in *Proceedingsof the International Conference on Advances in Computing,Communications and Informatics (ICACCI ’13)*, pp. 343–347, IEEE, Mysore, India, August 2013.

[29] G.Dougherty, *Digital Image Processing forMedical Applications*, Cambridge University Press, 2009.

[30] R. C. Gessner, C. B. Frederick, F. S. Foster, and P. A. Dayton, “Acoustic angiography: a new imaging modality for assessing microvasculature architecture,” *International Journal of BiomedicalImaging*, vol. 2013,Article ID 936593, 9 pages, 2013.

[31] Janarthanan Y, Balajee J.M, and SrinivasaRaghavaS."Content based video retrieval and analysis usingimage processing: A review."International Journal ofPharmacy and Technology 8, no.4 (2016): 5042-5048.

[32] Jeyakumar, Balajee, MA SaleemDurai, and DaphneLopez. "Case Studies in Amalgamation of DeepLearning and Big Data." In HCI Challenges andPrivacy Preservation in Big Data Security, pp. 159-174. IGI Global, 2018.

[33] Kamalakannan, S. "G., Balajee, J., Srinivasa Raghavan., “Superior content-based video retrievalsystem according to query image”." InternationalJournal of Applied Engineering Research 10, no. 3(2015): 7951-7957.

[34] Sethumadahavi R Balajee J “Big Data Deep Learningin Healthcare for Electronic Health Records,”International Scientific Research OrganizationJournal, vol. 2, Issue 2, pp. 31–35, Jul. 2017.

[35] Ushapreethi P, BalajeeJeyakumar and BalaKrishnanP, Action Recongnition in Video SurvillanceUsingHipi and Map Reducing Model, International Journalof Mechanical Engineering and Technology 8(11),2017,pp. 368–375.