Research Scenario of Bio Informatics in Big Data Approach

S.Jafar Ali Ibrahim[1], Dr.M.Thangamani[2], D. Sarathkumar[3]

[1]Doctoral Research Fellow, School of Information and Communication Engineering, Anna University, Chennai, Tamil Nadu, India

[2]Assistant Professor, Department of Computer Technology, Kongu Engineering College, Perundurai, Tamil Nadu, India

[3]Assistant Professor, Department of Electrical & Electronics Engineering, Kongu Engineering College, Perundurai, Tamil Nadu, India

***Email:** dsarathkumareee@gmail.com*

DOI: http://doi.org/10.5281/zenodo.2596987

Abstract

Big Data can unify all patient related data to get a 360-degree view of the patient to analyze and predict outcomes. This investigation examines the concepts and characteristics of Big Data, concepts about Translational Bio Informatics and some public available big data repositories and major issues of big data. This issue covers the area of medical and healthcare applications and its opportunities.

***Keywords:** Big Data, Bio Informatics, Drug Discovery, Computational Intelligence Methods, Health Informatics, Health care data mining*

Big Data Concepts

Big data is a blanket term for the non-traditional strategies and technologies needed to gather, organize, process, and gather insights from large datasets. Characteristics of big data can be described us 6 V's, that are following Volume, Velocity, Variety, Value, Variability and Veracity [1, 2, 3]

Volume

It refers to as terabytes, petabytes, and zettabytes of data. This focus on near instant feedback has driven many big data practitioners away from a batch-oriented approach and closer to a real-time streaming system. Data is constantly being added, massaged, processed, and analyzed in order to keep up with the influx of new information and to surface valuable information early when it is most relevant.

Variety

While more traditional data processing systems might expect data to enter the pipeline already labeled, formatted, and organized, big data systems usually accept and store data closer to its raw state.

Big data life cycle looks like

So how is data really handled when managing with a big data framework? While ideas to exertion differ, there are some populace in the scenario and software that we can discuss for the most part. While the means exhibited underneath won't not be valid in all cases, they are broadly utilized.

The general tier of task embroiled with big data processing is:

- Ingesting data into the system
- Persisting the data in storage
- Computing and Analyzing data
- Visualizing the results

In Big data technology, we will take a moment to talk about clustered computing, an important strategy employed by most big data solutions.

CLUSTERED COMPUTING

Resource Pooling: Combining the available storage space to hold data is a clear benefit, but CPU and memory pooling is also extremely important.

High Availability: Clusters can provide varying levels of fault tolerance and availability guarantees to prevent hardware or software failures from affecting access to data and processing.

Easy Scalability: Clusters make it easy to scale horizontally by adding additional machines to the group.
There is often noisy data or false information in big data. The focus of Big Data is on correlations, not causality [4].

CATEGORIES OF MEDICAL BIG DATA

Data in healthcare can be categorized as follows.

Genomic Data

Such data are gathered by a bioinformatics system or genomic data processing software. Data sequencing analysis techniques and variation analysis are common processes performed on genomic data. The aim of genomic data analysis is to determine the functions of specific genes. It refers to genotyping, gene expression and DNA sequence [6, 7].

Clinical Data

A term defined in the context of a clinical t rial for data pertaining to the health status of a patient or subject [8]. About 80% of this type data are unstructured documents, images and clinical or transcribed notes [9] Structured data (e.g., laboratory data, structured EMR/HER)

Behaviour Data and Patient Sentiment Data

Behavioural data refers to information produced as a result of actions, typically commercial behaviour using a range of devices connected to the Internet, such as a PC, tablet, or Smartphone. Behavioural data tracks the sites visited, the apps downloaded, or the games played. • Web and social media data Search engines, Internet consumer use and networking sites (Facebook, Twitter, Linkedin, blog, health plan websites and smartphone, etc.) [10]

Clinical reference and health publication data

It refers to reference data for clinical, claim, and business data to enable interoperability, drive compliance, and improve operational efficiencies.
Text-based publications (journals articles, clinical research and medical reference material) and clinical text-based reference practice guidelines and health product (e.g., drug information) data [7, 12].

Administrative, Business and External Data

- Insurance claims and related financial data, billing and scheduling [10]
- Biometric data: Fingerprints, handwriting and iris scans, etc
- Other Important Data
- Device data, adverse events and patient feedback, etc. [9]
- The content from portal or Personal Health Records (PHR) messaging (such as e-mails) between the patient and the provider team; the data generated in the PHR Ingesting data into the system
- Persisting the data in storage
- Computing and Analyzing data
- Visualizing the results

Big data in Health Informatics:

However, the scope of this study will be research that uses data mining in order to answer questions throughout the various levels of health[13].

The scope of data used by the subfield TBI, on the other hand, exploits data from each of these levels, from the molecular level to entire populations [14].

BIG DATA AND DRUG DISCOVERY

In today drug discovery environment, Big Data plays a vital role due to its 5 V concepts. These databases provide information about the drugs, their adverse

reactions, 1chemical formula, information about metabolic pathways, drug targets, disease for which a particular drug is used etc. None of the existing pharmacogenomic databases carry the complete integrated information and hence there is a need to develop a database which integrates data from all the widely used databases [38].
Integrating big data analytics and validating drugs in silico has the potential to improve the cost-effectiveness of the drug development pipeline. Big data–driven strategies are being increasingly used to address these challenges. Computational prediction of drug toxicity and pharmacodynamic/pharmacokinetic properties, based on integration of multiple data types, helps prioritize compounds for in vivo and human testing, potentially reducing costs[39].

DRUG DISCOVERY RELATED BIG DATA SOURCES

Data sets and resources available on Related to drug discovery are scattered in various databases and online resources and most of these databases are interlinked based on the information they carry. Some of these databases include PharmGKB [40], DrugBank [41], CTD [42], Reactome [43], KEGG [46], STITCH [47], PACdb [48], dbGaP [49] IGVdb, PGP [50]. Brief explanation of the databases are given in the following section and also tabulated in table 2.

PharmGKB

PharmGKB is a pharmocogenomics database that carries all the clinical information along with the dosage guidelines, gene-drug associations and genotype phenotype relationships. It also has information about Variant Annotations, Clinical Annotations and Very Important Pharmacogene (VIP) summaries, drug-centered pathways.

Drug Bank

Drug Bank database is the open resource for drug, drug targets, and chemo informatics. It contains 11,067 drug entries including 2,525 approved small molecule drugs, 960 approved biotech (protein/peptide) drugs, 112 nutraceuticals and over 5,125 experimental drugs. Additionally, 4,924 non-redundant protein (i.e. drug target/enzyme/transporter/carrier) sequences are linked to these drug entries. Each Drug Card entry contains more than 200 data fields with half of the information being devoted to drug/chemical data and the other half devoted to drug target or protein data.

CTD

The whole database is categorized in to 11 types:
Chemicals, genes, chemical-gene/protein interactions, diseases, gene-disease associations, chemical-disease associations, references, organisms, gene ontology, pathways and exposures.

Reactome

It has cross-referenced to several other databases such as Ensembl [44] and UniProt. The pathways within the database especially those pertaining to those in humans may be used for research and analysis, pathways modelling, systems biology as well as pharmacogenomics applications to analyze effects of drug pathway alterations on drug response and phenotypes [45].

KEGG

It is an integrated resource of systems information (KEGG Pathways, KEGG Brite, KEGG Module, KEGG Disease, KEGG Drug and KEGG Environ), genomics information (KEGG Orthology, KEGG Genes, KEGG Genome, KEGG DGenes and KEGG SSDB) and chemical information (KEGG Compounds, KEGG Glycans, KEGG Reaction, KEGG RPair, KEGG RClass and KEGG Enzyme).

STITCH

STITCH (Search Tool for Interacting Chemicals) is a database of known and predicted interactions between chemicals and proteins. The interactions include direct (physical) and indirect (functional) associations; they stem from computational prediction, from knowledge transfer between organisms, and from interactions aggregated from other (primary) databases. It also includes data on interactions between 210,914 small molecules and 9’643’763 proteins from 2'031 organisms

Other databases

dpGaP (Database of Genotypes and Phenotypes) is database of genotype-phenotype association studies, genome-wide association studies, as well as associations between genotype and non-clinical traits. It was developed to archive and distribute the data and results from studies that have investigated the interaction of genotype and phenotype in Humans.

PACdb (Pharamacogenomics and Cell database) contains information on the relationships between SNPs, gene expression and cellular sensitivity to drugs analyzed in cell-based models. It is a Pharmacogenetics-Cell line database for use as a central repository of pharmacology-related phenotypes that integrates genotypic, gene expression, and pharmacological data obtained via lymphoblastoid cell lines. 90 YRI LCLs as well as ExiqonmiRNA baseline data from 60 unrelated CEU and 60 unrelated YRI have been deposited in the PACdb database.

IGVd (Indian Genome Variation database) contains information about SNP, CNVs in over
1000 genes of biomedical important metabolic and genetic networks and also genes of pharmacogenetic relevance [51].

There are many other biological databases such as Uniprot, GO, GenBank, PDB have cross-reference to above databases whose information may serve as essential source for drug and it related studies.

Table 1: *Levels of Data*

Sections	Data level(s) Used	Subsections	Question level(s) answered	Questions to be answered
Using Micro Level Data – Molecules	Molecular	Using Gene Expression Data to Make Clinical Predictions	Clinical	1. What sub-type of cancer does a patient have? [18] 2. Will a patient have a relapse of cancer? [19]
Using Tissue Level Data	Tissue	Creating a Connectivity Map of the Brain Using Brain Images	Human-Scale Biology	Can a full connectivity map of the brain be made [20,21]?
	Patient	Using MRI Data for Clinical Prediction	Clinical	Do particular areas of the brain correlate to clinical events? [22]
Using Patient Level Data	Patient	Prediction of ICU Readmission and Mortality Rate	Clinical	1. Should a patient be released from the ICU, or would they benefit from a longer stay?[23-25] 2. What is the 5 year expectancy of a patient over the age of 50? [26]
		Real-Time Predictions Using Data Streams		1. What ailment does a patient have (real-time prediction) [27,28] 2. Is an infant experiencing a cardiorespiratory spell (real-time)? [29]
Using Population Level Data – Social Media	Population	Using Message Board Data to Help Patients Obtain Medical Information	Clinical	Can message post data be used for dispersing clinically reliable information? [30,31]
		Tracking Epidemics Using Search Query Data	Epidemic-Scale	Can search query data be used to accurately track epidemics throughout a population? [32,33]
		Tracking Epidemics Using Twitter Post Data	Epidemic-Scale	Can Twitter post data be used to accurately track epidemics throughout a population?[34,35]

Table 2: *Some Bio Informatics related Big Data Resources Which is publicly available*

Category	**Name**	**Description**	**URL**
Literature mining	PolySearch 2.0	Web-based text mining tool	http://polysearch.cs.ualberta.ca
Machine learning	Weka	Extensive library of machine learning algorithms with a user-friendly interface	http://www.cs.waikato.ac.nz/ml/weka/
Cheminformatics	Drug Bank Database	Database of drug chemical, structural, pharmacological, and target information	http://www.drugbank.ca
	PubChem	Comprehensive database of structural, pharmacological, and biochemical activity data	https://pubchem.ncbi.nlm.nih.gov/
	Protein Data Bank	Repository of protein structural data	http://www.wwpdb.org
	admetSAR	Web tool predicting pharmacological and toxicology parameters based on chemical structures	http://lmmd.ecust.edu.cn:8000/
	The Drug Gene Interaction Database (DGIdb)	Database of known drug-gene connections for selected genes	http://dgidb.genome.wustl.edu/
	SIDER	Database of drug adverse effects	http://sideeffects.embl.de/
	Library of Integrated Cellular Signatures (LINCS)	Database of functional cellular responses to genetic and pharmacological perturbations measured in multiple types of biomolecules (eg,transcriptome and kinome)	http://lincsportal.ccs.miami.edu/datasets/
	ChemBank	Database/knowledge base of high-throughput compound screens and other small molecule–related information	http://chembank.broadinstitute.org/
Molecular pathway knowledgebase/ analysis tool	DAVID	Searchable/downloadable database of molecular pathway knowledge base	https://david.ncifcrf.gov/
	NDEx	Biological network knowledge base	http://www.home.ndexbio.org/
	Molecular Signatures Database (MSigDb)	Repository of molecular signatures from curated databases, publications, and research studies	http://www.broadinstitute.org/msigdb
Omics data repositories	Gene Expression Omnibus	Repository of raw and processed omics data	http://www.ncbi.nlm.nih.gov/geo/
	Sequence Read Archive	Repository of sequencing data	http://www.ncbi.nlm.nih.gov/sra
	Array Express	Repository of raw and processed omics data	https://www.ebi.ac.uk/arrayexpress/
	The Cancer Genome Atlas	Repository of genomic, proteomic, histological, and clinical data for a wide variety of cancers	https://tcga-data.nci.nih.gov/tcga/ tcgaHome2.jsp

CONCLUSION

Big data is a broad, rapidly evolving topic. This survey discussed a number of recent studies being done within the most popular sub branches of Health Informatics, using Big Data from all accessible levels of human existence to answer questions throughout all levels. Analyzing Big Data of this scope has only been possible extremely recently, due to the increasing capability of both computational resources and the algorithms which take advantage of these resources. Research on using these tools and techniques for Health Informatics is critical, because this domain requires a great deal of testing and confirmation before new techniques can be applied for making real world decisions across all levels. The fact that computational power has reached the ability to handle Big Data through efficient algorithms. The use of Big Data provides advantages to Health Informatics by allowing for more tests cases or more features for research, leading to both quicker validations of studies.

REFERENCES

1. Eaton, C., D. Deroos, T. Deutsch, G. Lapis and P. Zikopoulos, 2012. Understanding big data. McGraw-Hill Companies.
2. O'Reilly Radar Team, 2012. Planning for big data. O'Reilly.
3. Zikopoulos, P., C. Eaton, D. de Roos, 2012. Understanding big data: Analytics for enterprise class hadoop and streaming data. McGraw-Hill, New York.
4. Bottles, K. and E. Begoli, 2014. Understanding the pros and cons of big data analytics. Physician Exec., 40: 6-12
5. Zaslavsky, A., C. Perera and D. Georgakopoulos, 2012. Sensing as a service and big data. Proceedings of the International Conference on Advances in Cloud Computing (ACC' 12), Bangalore, India, pp: 1-8.
6. Chen, H.C., R.H.L. Chiang and V.C. Storey, 2012. Business intelligence and analytics: From big data to big impact. MIS Q., 36: 1165-1188.
7. Priyanka, K. and N. Kulennavar, 2014. A survey on big data analytics in health care. Int. J. Comput. Sci. Inform. Technologies, 5: 5865-5868.
8. *Segen's Medical Dictionary*. S.v. "clinical data." Retrieved April 13 2018 from https://medicaldictionary.thefreedictionary.com/clinical+data
9. Yang, S., M. Njoku and C.F. Mackenzie, 2014. 'Big data' approaches to trauma outcome prediction and autonomous resuscitation. Brit. J. Hospital Med., 75: 637-641. DOI: 10.12968/hmed.2014.75.11.637.
10. Terry, N.P., 2013. Protecting patient privacy in the age of big data. UMKC Law Rev., 81: 385-415.
11. Shrestha, R.B., 2014. Big data and cloud computing. Applied Radiology.
12. Miller, K., 2012. Big data analytics in biomedical research. Biomedical Computation Review.
13. Herland *et al.*: A review of data mining using big data in health informatics. *Journal of Big Data*2014 **1**:2. doi:10.1186/2196-1115-1-2
14. Chen J, Qian F, Yan W, Shen B (2013) Translational biomedical informatics in the cloud: present and future. BioMed Res Int2013:8.[http://dx.doi.org/10.1155/2013/658925]
15. McDonald E, Brown CT (2013) khmer: Working with big data in Bioinformatics. CoRR abs/1303.2223: 1–18
16. Beltrame, F. and Koslow, S. H. (1999). Neuroinformatics as a mega science issue. *IEEE Transactions on Information Technology in Biomedicine*, 3(3):239-240. PMID: 10719488.
17. Luscombe, N. M., Greenbaum, D., and Gerstein, M.(2001). What is bioinformatics? a proposed definition and overview of the field. *Method. Inform. Med.*, 40(4):346-258. PMID: 11552348.
18. Haferlach T, Kohlmann A, Wieczorek L, Basso G, Kronnie GT, Béné MC, De Vos J, Hernández JM, Hofmann WK, Mills KI, Gilkes A, Chiaretti S, Shurtleff SA, Kipps TJ, Rassenti LZ, Yeoh AE, Papenhausen PR, Wm Liu, Williams PM, For (2010) Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: report from the international microarray innovations in leukemia study group. J Clin Oncol 28(15): 2529–

2537.[http://jco.ascopubs.org/content/28/15/2529.abstract]

19. Salazar R, Roepman P, Capella G, Moreno V, Simon I, Dreezen C, Lopez-Doriga A, Santos C, Marijnen C, Westerga J, Bruin S, Kerr D, Kuppen P, van de Velde C, Morreau H, Van Velthuysen L, Glas AM, Van't Veer LJ, Tollenaar R (2011)Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer. J Clin Oncol29: 17–24. [http://jco.ascopubs.org/content/29/1/17.abstract]
20. Annese J (2012) The importance of combining MRI and large-scale digital histology in neuroimaging studies of brain connectivity and disease. Front Neuroinform6:13.[http://europepmc.org/abstract/MED/22536182]
21. Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K (2013) The WU-Minn human connectome project: an overview. NeuroImage 80(0): 62–79. [http://www.sciencedirect.com/science/article/pii/ S1053811913005351]. [Mapping the Connectome]
22. Yoshida H, Kawaguchi A, Tsuruya K (2013) Radial basis function-sparse partial least squares for application to brain imaging data. Comput Math Methods Med 2013: 7. [http://dx.doi.org/10.1155/2013/591032]
23. Campbell AJ, Cook JA, Adey G, Cuthbertson BH (2008) Predicting death and readmission after intensive care discharge. British J Anaesth 100(5): 656–662. [http://europepmc.org/abstract/MED/18385264]
24. Fialho AS, Cismondi F, Vieira SM, Reti SR, Sousa JMC, Finkelstein SN (2012) Data mining using clinical physiology at discharge to predict ICU readmissions. Expert Syst Appl 39(18): 13158–13165. [http://www.sciencedirect.com/science/article/pii/S0957417412008020]
25. Ouanes I, Schwebel C, Franais A, Bruel C, Philippart F, Vesin A, Soufir L, Adrie C, Garrouste-Orgeas M, Timsit JF, Misset B (2012) A model to predict short-term death or readmission after intensive care unit discharge. J Crit Care 27(4): 422.e1–422.e9. [http://www.sciencedirect.com/science/article/pii/S0883944111003790]
26. Mathias JS, Agrawal A, Feinglass J, Cooper AJ, Baker DW, Choudhary A (2013) Development of a 5 year life expectancy index in older adults using predictive mining of electronic health record data. J Am Med Inform Assoc 20(e1): e118–e124. [http://jamia.bmj.com/content/20/e1/e118.abstract]
27. Ballard C, Foster K, Frenkiel A, Gedik B, Koranda MP, Nathan S, Rajan D, Rea R, Spicer M, Williams B, Zoubov VN (2011) IBM Infosphere Streams: Assembling Continuous Insight in the Information Revolution. [http://www.redbooks.ibm.com/abstracts/sg.pages=247970html]
28. Zhang Y, Fong S, Fiaidhi J, Mohammed S (2012) Real-time clinical decision support system with data stream mining.J Biomed Biotechnol 2012: 8. [http://dx.doi.org/10.1155/2012/580186]
29. Thommandram A, Pugh JE, Eklund JM, McGregor C, James AG (2013) Classifying neonatal spells using real-time temporal analysis of physiological data streams: Algorithm development In: IEEE Point-of-Care Healthcare Technologies (PHT 2013). IEEE, based in New York, USA, Bangalore, India, pp 240–243
30. Ashish N, Biswas A, Das S, Nag S, Pratap R (2012) The Abzooba smart health informatics platform (SHIP)™–frompatient experiences to big data to insights. CoRR abs/1203.3764: 1–3

31. Rolia J, Yao W, Basu S, Lee WN, Singhal S, Kumar A, Sabella S (2013) Tell me what i don't know - making the most of social health forums. Tech. Rep: HPL-2013–43. Hewlett Packard Labs [https://www.hpl.hp.com/techreports/2013/ HPL-2013-43.pdf]
32. Yuan Q, Nsoesie EO, Lv B, Peng G, Chunara R, Brownstein JS (2013) Monitoring influenza epidemics in China with search query from Baidu. PLoS ONE 8(5): e64323. [doi: 10.1371/journal.pone.0064323]
33. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L (2009) Detecting influenza epidemics using search engine query data. Nature 457(7232): 1012–1014. [http://dx.doi.org/10.1038/nature07634]
34. Achrekar H, Gandhe A, Lazarus R, Yu SH, Liu B (2012) Twitter improves seasonal influenza prediction In: International Conference on Health Informatics (HEALTHINF'12). Nature Publishing Group, based in London, UK, Vilamoura, Portugal, pp 61–70
35. Signorini A, Segre AM, Polgreen PM (2011) The use of twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS ONE 6(5): e19467. doi:10.1371/journal.pone.0019467
36. Bennett C, Doub T (2011) Data mining and electronic health records: selecting optimal clinical treatments in practice. CoRR abs/1112: 1668
37. Yasnoff WA, O'Carroll PW, Koo D, Linkins RW, Kilbourne EM. Public health informatics: improving and transforming public health in the information age. J Public Health Manag Pract 2000; 6:67–75.
38. Kumar, Pavan & Ch, Janaki & Neeharika, N & Saluja, Payal & Mangala, Natampalli & B.B, Prahlada Rao. (2015). Information gateway for integrated pharmacogenomics data-IGIPD. Proceedings - 2014 IEEE International Conference on Big Data, IEEE Big Data 2014. 1-9. 10.1109/BigData.2014.7004385.
39. Wang Y, Xing J, Xu Y, et al. In silico ADME/T modelling for rational drug design. Q Rev Biophys 2015; 48:488–515.
40. M. Whirl-Carrillo, E.M. McDonagh, J. M. Hebert, L. Gong, K. Sangkuhl, C.F. Thorn, R.B. Altman and T.E. Klein. "Pharmacogenomics Knowledge for Personalized Medicine"*Clinical Pharmacology & Therapeutics* (2012) 92(4): 414-417.
41. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, Chin L, Cummings R, Le D, Pon A, Knox C, Wilson M. Drug Bank 5.0: a major update to the Drug Bank database for 2018. Nucleic Acids Res. 2017 Nov 8. doi: 10.1093/nar/gkx1037.
42. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, King BL, McMorran R, Wiegers J, Wiegers TC, Mattingly CJ. The Comparative Toxicogenomics Database: update 2017. Nucleic Acids Res. 2016 Sep 19;[Epub ahead of print] PMID:27651457
43. The Reactome Pathway Knowledgebase. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B, Korninger F, May B, Milacic M, Roca CD, Rothfels K, Sevilla C, Shamovsky V, Shorser S, Varusai T, Viteri G, Weiser J, Wu G, Stein L, Hermjakob H, D'Eustachio P. Nucleic Acids Res. 2018 Jan 4;46(D1):D649-D655.doi:10.1093/nar/gkx1132.PMID: 29145629
44. Daniel R. Zerbino. Et al, **Ensembl 2018.** PubMed PMID: 29155950. doi:10.1093/nar/gkx1098.
45. Ayesha Pasha, Vinod Scaria, "Pharmacogenomics in the Era of

Personal Genomics: A Quick Guide to Online Resources and Tools", Omics for Personalized Medicine, pp. 187-211, 2013
46. Kanehisa, Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K.; KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 45, D353-D361 (2017).
47. Kuhn, Michael et al. “STITCH: Interaction Networks of Chemicals and Proteins.” Nucleic Acids Research 36.Database issue (2008): D684–D688. PMC. Web. 26 Apr. 2018.
48. Gamazon, Eric R. et al. “PACdb: A Database for Cell-Based Pharmacogenomics.” Pharmacogenetics and genomics 20.4 (2010): 269–273. PMC. Web. 26 Apr. 2018.
49. Mailman MD et al, 'The NCBI dbGaP database of genotypes and phenotypes", Nat Genet, vol. 39, no. 10, pp. 1181–1186, 2007.
50. PGP-UK: a research and citizen science hybrid project in support of personalized medicine. Stephan Beck et al bioRxiv 288829; doi: https://doi.org/10.1101/288829
51. The Indian Genome Variation Consortium Hum Genet (2005) 118: 1. https://doi.org/10.1007/s00439-005-0009-9
52. Ibrahim, S. Jafar Ali, and M. Thangamani. "Prediction of Novel Drugs and Diseases for Hepatocellular Carcinoma Based on Multi-Source Simulated Annealing Based Random Walk." Journal of medical systems 42, no. 10 (2018): 188.
53. *telemedicine and applications* 2018 (2018).

Bibliography

Jafar Ali Ibrahim. S has completed his Bachelor of Technology in Information Technology from Syed Ammal Engineering College, Ramanathapuram, affiliated to Anna University, Chennai, Tamilnadu, India, and Master of Technology degree in Computer Systems and Networks from Dr. MGR Educational and Research Institute, Chennai, Tamil Nadu, India. He possesses nearly 5 years of Industrial experience in the field of IT Security, Bio Metric Deployment, IT Administration, Network Security, Web Services and Security, Service Oriented Architecture environment etc. He has presented the paper in 11 reputed International conferences and published twelve papers in reputed journals. His research interest includes Cloud Computing, Clinical Research, and Medical Informatics, Internet of Things, Machine Learning, Ontology Development, Big Data, and Data mining. He visited the countries like Japan, Malaysia, Singapore, Srilanka, and Gulf countries for his research activities. Right Now he is doing his Ph.D. as a full-time Doctoral Research Fellow in the area of Translational Clinical Informatics at Anna University, Chennai, Tamilnadu, India.

Dr. M. Thangamani possesses nearly 23 years of experience in research, teaching, consulting and practical application development to solve real-world business problems using analytics. Her research expertise covers Medical data mining, machine learning, cloud computing, big data, fuzzy, soft computing, ontology development, web services, and open source software. She has published nearly 80 articles in refereed, indexed, SCI Journals, books, and book chapters and presented over 67 papers in national and international conferences in the above field. She has delivered more than 60 Guest Lectures in reputed engineering colleges and reputed industries on various topics. She has got best paper awards from various education-related social activities in India and Abroad. She has received the many National and International Awards. She continues to actively serve the academic and research communities and presently guiding nine Ph.D. Scholars under Anna University. She is on the editorial board and reviewing committee of leading research SCI journals. She has on the program committee of top international data mining and soft computing conferences in various countries. She is also Board member in Taylor & Francis Group and seasonal reviewer in IEEE Transaction on Fuzzy System, the international journal of advances in Fuzzy System and Applied mathematics and information journals. She has an organizing chair and keynotes speaker in international conferences in India and countries like California, Dubai, Malaysia.


D.Sarathkumar, M.E., is working as an Assistant Professor in Department of Electrical and Electronics Engineering, Kongu Engineering College, Erode from June 2014 onwards. He graduated B.E (EEE) from Sri Ramakrishna Institute of technology in the year 2012 and M.E (Power Systems Engineering) from Velammal Engineering College in the year 2014 under Anna University, Chennai. He has published 6 papers in International journals, 3 papers is national journals and presented 5 papers in international conferences and 3 papers in national conferences. He organized more than 25 workshops, seminars and short term training programs in various fields. His Areas of interests include Data Communication Networks, Soft Bio Informatics in Big data approach, Internet of things (IOT), Power System Analysis and Stability, FACTS, Energy Utilization and conservation etc. His research interests include Big data Analytics, IOT in Healthcare, Power System Stability and power Quality Improvement using various Custom Power Devices.

Cite this Article as:
S.Jafar Ali Ibrahim, Dr.M.Thangamani, & D. Sarathkumar. (2019). Research Scenario of Bio Informatics in Big Data Approach. Journal of Electronics and Communication Systems, 4(1), 18–27. http://doi.org/10.5281/zenodo.2596987

