Modern Education
and Computer Science
PRESS

Managing Data Diversity on the Internet of Medical Things (IoMT)

Iram Mehmood
GC Woman University Sialkot / Department of Computer Science & Information Technology, Pakistan
E-mail: IramKhalid.cs27@gmail.com

Sidra Anwar
Memorial University of Newfoundland, St. John's, Canada
E-mail: sidraa@mun.ca

AneezaDilawar
GC Woman University Sialkot / Department of Computer Science & Information Technology, Pakistan
E-mail: aneezadilawar786@gmail.com

IsmaZulfiqar
GC Woman University Sialkot / Department of Computer Science & Information Technology, Pakistan
E-mail: csismazulfiqar@gmail.com

Raja Manzar Abbas
University of Limerick, Limerick/ Department of Computer Science, Ireland
E-mail: manzar.abbas@ul.ie

Abstract: In the healthcare industry, the Internet of Medical Services (IOMT) plays a vital role throughout the increasing performance, reliability, and efficiency of an electronic device. Healthcare is also characterized as being complicated due to its highly diverse and large number of shareholders. Data diversity refers to the continuum of various types of elements in the data. The integration of data is difficult where different sources can adopt different identification for the same entity, but there is no explicit connection. Researches are contributing to a digitized Health care system through interconnections available medical resources and health care services. This Research presents the contribution of IoT to people in the field of Healthcare, highlighting the issues in different data integration, analysis of the existing algorithms and models, applications, and future challenges of IoT in terms of healthcare medical services. Big data analytics that incorporates millions of fragmented, organized, and unstructured sources of data will play a key role in how health care will be delivered in the future.

Index Terms: Big data, internet of medical things, diversity.

1. Introduction

Ref. [1]National and various public participants have moved towards consistency by generating decades of relevant, accessible, and enforceable information for the system as a whole. These will increase the information volatility to the tipping purpose of the business. This data can be a type of "big data," which is not only known by its sheer scale as well as for its scalability, variety, punctuality. Ref. [2] The term variety is that the quality of being different and numerous.

Data variety, however, plays a significant role in process concerns. Ref. [1,2] Though the machine power of process systems has been steadily increasing, and the information diversity is rising quicker.

In the attention business, the Internet of Medical Things (IOMT) plays a major role in increasing the efficiency, reliability and profitability of electronic equipment.Ref. [3] The Internet of things is not a concept of substitution, but it is a significant issue in the universe.

Ref. [4] Healthcare is additionally represented to be advanced as a result of extraordinarily numerous and enormous variety of their shareholders. Additionally referred to as tending IOT, the web of Medical Things contains the medical-equipment and applications connected to managing IT systems via the net. Wi-Fi enabled devices to facilitate

machine-to-machine communication and link to cloud platforms for knowledge storage.

Ref. [3] Consecutive generation of technological world is the Things Web (IoT) that links equipment, sensor systems, electronics, vehicles and different "things." The term web of Medical Things (IOMT), care application of IOT technology, includes a network of connected devices that sense essential knowledge in real-time.

Ref. [1,3] Healthcare trade is historically a trade of information, that is understood for continuous generation of information from all its numerous activities. This information area unit characteristically voluminous, heterogeneous, incredibly messed up and various, this is often as a result of health care trade itself could be an advanced trade.

Ref. [6] Healthcare knowledge is a 'big-data'; this is often not solely thanks to its size, however additionally thanks to the speed at that is being generated, its selection, and un-structured-ness.

Ref. [4, 6] diversity of information refers to the spectrum of various part sorts at intervals the information. Knowledge is often considerably numerous, or instance, mobile devices like sensible phones square measure ready to manufacture varied kinds of data, e.g., audio, video or transmission knowledge, text data, location knowledge, temporal knowledge, rotating knowledge.

Ref. [5] Different information sorts typically need different handling and impose numerous need soon allocation capability, speed of the process, and alternative problems. Information diversity adds richness; however, it additionally adds complexities to operation and maintenance. Organizations should account for numerous conditions and needs associated with information diversity. There are several issues regarding data diversity in hand.

The data in IOMT can pose various formats on their structures with different constraints. In a BDAaaS case, where the addition of a new complex data set has to be done automatically, the issue of data diversity can be compounded even further. Nevertheless, from the essence of the information, it includes, there are many other problems at hand concerning the study and management of big data diversity. The datasets are in large sizes and contain heterogeneous laws, and patterns show significant variation in characteristics.

The purpose of this Research is to highlight the data diversity issues during data integration (Section A), and present some algorithms and models using for big data diversity in(Section III & IV) .

1.1. Issues in Diverse Data Integration

An emphasis on how records that belong to the same entity can be identified and connected within one or through multiple data sets and discussed as follows:

- A critical stage in data integration is the mission, which goes by names such as data linkage, record linking, data matching, or object resolution.
- Schema mapping and data fusion are two further steps.
- Duplicate identification when you need to classify information of the same entity in a single collection of data.
- The lack of a specific entity identifier across the data sets is a major issue when connecting records from various sources.

Ref. [32] In more recent times, informatics analysts have addressed data integration from a conventional binary supervised classification viewpoint (with the two classes matching and non-matching, but without possible matches) or through a clustering viewpoint. Ref. [33] One problem for both conventional probabilistic data integration and modern pairly supervised identification techniques is that they often lead to non-transitivity tests. The aforementioned problem can be fixed by accepting the classification of the data linkage as a clustering question where it is intended to combine all records relating to the same entity under one cluster.

2. Literature Review

2.1. Background

Ref. [31] For the primary period of 1997, the word "big data" used by Michael Cox Associate in Nursing David Ell sworth in an excessively article conferred on laptop systems through an IEEE symposium to elucidate the mental image with information as well as the hurdles it presented.

From 1986 to 2007, worldwide capabilities of technical information storing, computing, scanning as well as interactions were caterpillar-monitored sixty stimulants or modern technologies.

Ref. [7] By even the top of only 1990s, the speedy Information technology developments also technological enhancements had allowed coevals of huge quantity belonging to knowledge; however, very small helpful information in relation.

Ref. [7] The duration of 2001 to 2008 used to be the biological process level for large information development.Huge information initial outline in terms of its magnitude, speed, and selection, when that it grew to be feasible to improve additional delicate package to meet the needs of coping with facts explosion consequently.

Ref. [7, 9] The Start of 2009, massive information analytics entered the innovative level. Not solely had machine-learning computing turns out to be an opening through innovation for enterprise brain, however additionally researchers

had been predicting the information organizational so included approaches that were getting ready to move in a piece of unstructured information to integrated information, and from a fixed fatal atmosphere to an omnipresent swarm primarily supported atmosphere.

Ref. [7] Programs and apps innovation like protractible nomenclature (XML) internet facilities, direction strategies, Hadoop another big data component as well as core modules working that centered on improve accessibility for site users, allow customers to method Brobdingnagian volume of data information through and inside entities collectively and also in live time.

Ref. [12] Google handled 20,000 TB of information a day in 2008. Ref. [14] Acamai tracks and evaluate75 million events a day to boost advertising. Walmart processes more than 1 million user transactions, producing data of more than 2.5 PB as a measurement.

Ref. [11] In 2010, the market place for huge knowledge was $3.2 billion, and this price is predicted to extend to $16.9 billion in 2015. Throughout 2012, 2.2 million TB of recent knowledge square measure generated day by day.

Ref. [13] In 2012 defines massive knowledge as massive volumes of variable knowledge in Health Care that need Advanced information collection, processing, delivery management and evaluation methods and technologies.

Ref. [13] At the same time, health care organizations were getting down to digitalize their patient's files and combination medical information to immense digital data sources. Such innovations create patient information accessibly, findable, workable, as well as enabled attention suppliers to follow more straightforward medication.

Ref. [2] Quantum theory incorporates the good innovative e-Science facilities above all thanks in it would like to suitable big data services to evaluating outcomes or even collect information beginning from the ECU organization for Nuclear Analysis (CERN) massive fundamental particle accelerator, that started taking information in 2009.

Ref. [1] The 2009 Government accountability guideline, still because to an ensuant force of the healthcare department below the Medical Information Strategy (MDS), area unit getting down to liberate information from agencies just like the institutes for Medical Services (IMS), the World health organization, and also institutes of sickness management.

Ref. [14, 1] The Health data Technology for Economy and Medical Health (HITEMH) Act, that was a part of the 2009 Yankee Restoration and Increased investment law, approved up to regarding $40 billion in incentive payments for suppliers to use EMRs, with the general objective of promoting acceptance to seventy to ninety percent of all suppliers by 2019.

Ref. [7, 13] As far as health care exchange is concerned, as of 2011, data on medical services had exceeded one hundred and fifty exabytes (1 EB = 1018 bytes) across the board, principally within the style of electronic medical records (Institute for Health Technology Transformation, 2013). Ref. [13] At this rate of growth, massive information for U.S. health care can shortly extend the zettabyte (1021 gigabytes) scale and directly when the yottabyte (1024 gigabytes).

Medical imaging generates huge amounts of knowledge with complicated options and wider dimensions.

The scheme of medicine massive knowledge includes many alternative levels of knowledge sources to form an expensive array of knowledge for researchers. Abundant of the information that area unit gives several chances or novel take exception in implementing different inquiries investigations.

2.2. Related Work

Ref. [16] The EU community had undertaken the potential of health telematics trade as early as 1985 when a cumulative endeavor was initiated to formalize a mid- to fast-term strategy for Europe in the formulation of data and technological developments to sustain and biomedical. That was afraid of the structure of an initiative called the European Program and Policy Co-operative for Bio-Informatics.

The Council of Ministers embraced a wildcat intervention on Gregorian calendar month four, 1988, with a twenty million EU community engagement. The program, Progress IP in medication, pursued the wildcat action in Gregorian calendar month 1991.

Ref. [16, 17] Telematics techniques and applications had been evolved in 1998 to facilitate in the expense-containment and better governance of the healthcare industry and to broaden the struggle towards the medical driver assistance business. For a considerable time, the older part of health telematics has proven the scenario of health promotion.

Ref. [23] Health Fidelity is the victimhood of natural language processing to demonstrate unorganized knowledge into organized knowledge suitable for device executives to properly manage requirements in the executives of the cash flow cycle, enforcement and analysis. The computer vision innovation of Medical Fidelity adherents challenging and advanced medical narratives and disintegrates important information to construct it in real-time on the sector.

Ref. [23] WellDoc is device controlled exploitation, a period of time-work that combines behavior and computerized medical communication, the mutual element of its section II medical equipment model for Food and Drug Administration, to help individuals handle chronic conditions such as gene expression disorder.WellDoc depicts a massive fashion of organized and unorganized patient-reported details regarding clinical and behavioral aspects of the sufferer's mobile or web-based application for medical exploitation.

Ref. [23] Next Bioorganizes enormous quantities of moderated and transcribed genomic expertise to sanction

buyers with firsthand knowledge to form unfeasible discoveries to their datasets.

Practice Fusion could be a public web-based medical education EMR application that also compiles patient information from numerous sites to enhance clinical Research and evaluation of public health. Ref. [23] The EMR system updates itself over time as technology improves, creating enhancements, including new customer-friendly mobile platforms. 1) Health Population Police Enquiry and Literacy 2) Post-Market Treatment Police Enquiry 3) Global health Assessment 4) Advancement of Treatment and Best Examine.

Ref. [27] CloudBurst can be a concurrent processing model that inhibits the method of order mapping. Rain parallels the brief-read visualization approach to improve the mapping of massive information and learning measurability.

Ref. [21] BITs square gauge interaction appliance and psychological involvement methods by using technology possibilities to manage behaviour, psychological characteristics, and affectional goals that endorse physical, interaction, and psychological state. BITs provide a variety of techniques, and also distant psychotherapy availability, internet-based and portable therapies, social networks, virtual reality, and people and vice. Such procedure square dimensions are currently in broad variable quantitative and maturity phases, involving dissects varying from fundamental production and evaluation to analyses of deployment.

Ref. [20] DistMap a framework on a Hadoop cluster for decentralized brief-read mapping. DistMap seeks to broaden the approval of numerous optimizer varieties to hide a broader range of applications for synthesizing.

Ref. [21] Myrnaa web-based computational pipeline that computes organic occurrence variants in gigantic RNA-seq datasets. Myrna endorses several RNA-seq assessment processes in an incorporated pipeline, along with reads orientation, optimization, and try to Identify in Nursingd applied math modeling.

Ref. [23] Explorys, Inc. has built a cloud-based application framework that compiles vast amounts of information from multiple diverse sources along with multiple stakeholders ' cash, operational and medical information. Then the data is structured and optimized to help consumers control populations by looking for trends of illness, diagnosis, and result.

Ref. [23] DNAnexus provides a cloud-based, community-based, supportive and ascending information proprietary technology with knowledge administrators, analysis and visualization of next-generation sequencing (NGS) to serve patients with betterment.

Table 1. Analysis of IoMT applications

Applications in Medical domain				
Sr #	Publisheddate	Author Name	Model/Application	Features
01	5-Feb-2014	Constantine Stephanidis	Health Telematics	facilitate in the expense-containment and better governance[16]
02	Oct-2012	Bonnie Feldman	Health Fidelity	demonstrate unstructured data into structured data suitable for deviceexecutives[23]
03	8-Apr -2009	Michael C. Schatz	CloudBurst	brief-read visualization approach inhibits method of order mapping[27]
04	Oct-2012	Tobi Skotnes	NextBio	organizes enormous quantities of genomic with 1st hand knowledge[23]
05	October 2012	Bonnie Feldman	Practice Fusion	compiles patient information from numerous sites to enhance clinical research[23]
06	Oct-2012	Ellen M.Martin	WellDoc	combines behavior and computerized medical communication depicts organized and unorganized patient-reported details[23]
07	Aug-2013	David C.Mohr Ph.D.	BITs square	Manage psychological behavior, internet-based therapies[27]
08	23-Aug-2013	Ram Vinay Pandey	DistMap	numerous optimizer varieties to hide applications for synthesizing[20]
09	October 2012	Bonnie Feldman	Explorys	Structure and optimize vast amounts of information from diverse sources[23]
10	11-Aug-2010	Ben Langmead	Myrna	computes organic occurrence variants in gigantic RNA-seq datasets[21]
11	October 2012	Ellen M. Martin	DNAnexus	analysis and visualization of next-generation sequencing[23]

"Table 1," Lists a number of IOT implementations from Year (2009-2014). There are five columns each representing some features, serial number, published Year, author name, application/model, and the characteristics from which we were researching. Many of the applications we 've been studying are from the healthcare systems. Most of the applications come from written research papers.

Moreover, several algorithms for recognizing different elements were proposed. The proposed algorithms also leverage different characteristics of the underlying issue of heterogeneity. We have also discussed two major classes of algorithms, namely greedy algorithms and graph algorithms.

A. Greedy algorithms

Ref. [8] Greedy methods are the most widely used for all concepts of variety, that is, distance-based, coverage-based, and novelty-based approaches, either with a view to optimizing diversity alone or with some notion of useful

ness or rating. Ref. [10] Generic Greedy algorithms which use two sets of items, the set I of all n elements available and the set S of diverse parts. The elements are moved from I to S recursively until |S| = k and |I|= n-k.

B. Neighborhood algorithms

Ref. [10] Neighborhood algorithms make up a particular case of greedy algorithms. Neighborhood algorithms were primarily used for distance-based models, where elements are picked in rounds, a similar strategy is also used, and each selected element contributes to the disqualification of its neighborhood.

C. Interchange algorithms

The interchange algorithms are another category of diversity algorithms used for all forms of diversity models. Ref. [15]. An interchange algorithm is used for a distance-based model balancing heterogeneity with utility. Ref. [18] An algorithm of interchange is used to attain diversity based on the coverage of structured data.

D. Graph algorithms

A graph-based algorithm, called GRASSHOPPER, was designed based on random walks. GRASSHOPPER uses a weighted graph that maps data item diversity as well as relevance. Ref. [20] Nodes are transcribed with a weight indicating their importance, while edges are transcribed with a weight indicating their neighboring nodes similarity. Ref. [21] The walker either travels to a neighboring state, with some possibility, as per to the similarity defined by edge weights or teleports according to usefulness to a random state. Ref. [21] The selected node is converted to an absorbing position at the end of the walk, and the walk is repeated. Another graph-based method is known as the Affinity Graph, which is formulated based on their resemblance but is determined asymmetrically.

Greedy Algorithm is best implemented in real time when a solution is needed. Greedy algorithm minimizes time while ensuring that an optimal solution is generated, hence use in a situation where less time is needed is more applicable.

3. Models

3.1. Distance-based measures

Ref. [24] Distance-based assessments of diversity depends on a pair-wise measure of the distance between the S components. Such a pair-wise measure of the diversity div(S) between two components of S is determined by the use of an aggregation function of the pair-wise distances between its components. Ref. [25] The most accepted definition of S diversity is either the average or the minimum. Defining an acceptable pair-wise distance measure is essential to successful diversification, and is strongly dependent on implementation.

3.2. Coverage-based measures

Ref. [26] Coverage-based diversity initiatives depend on a predefined number of factors, i.e., subjects, perceptions, or views, to exist. Coverage-based models vary in different ways, (a) what elements are to be addressed, and (b) how coverage is assessed.

3.3. Novelty-based measures

Ref. [24] The diversity initiatives centered on innovation describe diversity with regard to the elements seen in the past. Ref. [25] The key search and decision goal of novelty-based diversity is to reduce redundancy. For this, the elements are regularly picked one by one.

Table 2. Metrics of Frameworks

Techniques/Frameworks using in IoMT			
Sr#	Technique/Framework	Description	Limitation
01	Dist Map	framework on a Hadoop cluster for decentralized brief-read mapping	It can be difficult to distinguish between different shades in the maps.
02	Spark	General-purpose distributed data processing engine that is suitable for use in a wide range of circumstances. It can be used for batch processing.	No file management system, Expensive, small files issues and window criteria, no real-time processing
03	Storm	Free and open-source distributed real-time computation system, doing for real-time processing.	Update a running topology the only option is to kill the current topology and resubmit
04	Hadoop(Apache platform)	Hadoop framework is designed to scale up from a single machine to thousands of machines in a cluster	Small data, slow processing speed, batch processing only, Latency

"Table 2," Shows framework that is used in IOMT area. There are four columns, each representing some attribute, number of sequences, techniques / frame, description, and limitation. We selected these applications from different resources, based on the contribution of various researchers in the field of it.

Hadoop is the best of the techniques. It is a programming framework based on the open-source Java. In HDFS, Hadoop stores massive quantities of data in a distributed way. Hadoop is to split the data into several infrastructure Structures to process it. Process the data on a cluster of the node in parallel. Hadoop stack is versatile, scalable, economical, fault tolerance, and simple to use.

4. Discussion

The Internet of medical things is currently under discussion being a hot topic in the world. Ref. [13]. Because of the accelerated technology advances, big data, and the overall increase in information intensive activities, medical organizations face several difficulties: storage, Research, and the bottom line is divergence. Lack of common entity identifiers when data comes from diverse sources. There is no external linkage available for diverse data sets. Ref. [19] There is a need to warehouse and access various data modes for decisions that have pressed medical enterprises to look for better methods than traditional storage arrays and procedures. Many techniques in this area can be identified.

There is also discussed some existing applications for IoMT and a few algorithms to solve the problem of data diversity in IoMT. There are two major categories of algorithms discussed, namely, greedy and graph algorithms. Few models, for instance, distance-based, coverage-based, and novelty-based, were also discussed in this paper. Besides, various critical big data aspects are addressed, and the techniques are examined. One difficulty in the healthcare sector is that its acceptance of IT has typically slowed down behind the other areas, that is one of the primary reasons why it is hard to find instances. Most of the recorded instances emerged from vendors for this Research.

As a consequence, it is concluded that IoT based technologies and programs have turned the world into an alternative reality about which the 90's person was speaking. IoT helps the doctors and medical personnel to do their job with less energy and knowledge more reliably and effectively. Progression has been facilitated in parallel to involving empirical analysis of big data analytics.

6. Conclusion

The IOMT technological advances and its IoT applications provide varying dimensions and online services. Such apps have sometimes served millions of individuals. IoT systems generate the vast volume of computing in the healthcare setting. Through evaluating analytics, our work has given a better picture of how big data analytics can be used through health providers to optimize IT to acquire value for the business. Reviewing the effect on the medical entity, the quality of big data analytics techniques using a quantitative analysis process based on the available data may reflect different light.

Big data in Healthcare are challenging not only due to its scale but also due to the diversity of data types and the speed at which it must be managed. Therefore, there is a need for such a solution integrated with big data systems where it is possible to reduce vast variation in IOMT data and minimize the risks associated with incorrect analyzes.

With an increasing percentage of diverse and unstructured data, sophisticated analytical strategies are desperately needed to enable computers to identify interest in massive amounts of unstructured data and to conclude connections without any need for particular models programming directions. These techniques tackle certain concerns, incentives, obstacles, and the potential to use diverse data sets to expand precision and reduce costs and improve processing.

7. Future Work

As for future work, by collecting and analyzing actual data, further and improved exploration could be made. Therefore, in the big data analytics setting, we believe that future studies will take analytical personnel into account. Big data analytics that incorporates millions of fragmented, organized, and unstructured sources of data will play a key role in how health care will be delivered in the future.

Reference

[1] Groves, P., Kayyali, B., Knott, D., & Kuiken, S. V. (2016). The'big data' revolution in Healthcare: Accelerating value and innovation.

[2] Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information sciences*, *275*, 314-347.

[3] Joyia, G. J., Liaqat, R. M., Farooq, A., & Rehman, S. (2017). Internet of Medical Things (IOMT): applications, benefits and future challenges in healthcare domain. *J Commun*, *12*(4), 240-7.

[4] Taiwo, A. (2019). Big-Data Technology in Finnish Healthcare: Barriers and Possible Ways Out.

[5] Drosou, M., Jagadish, H. V., Pitoura, E., & Stoyanovich, J. (2017). Diversity in big data: A review. *Big data*, *5*(2), 73-84.

[6]   Tukkoji, C. D., & Seetharam, K. A Survey on Big Data Traffic Management with the Help of Cloud computing (for Biometric systems).

[7]   Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, *126*, 3-13.

[8]   Erkut, E., Ülküsal, Y., & Yenicerioğlu, O. (1994). A comparison of p-dispersion heuristics. *Computers & operations research*, *21*(10), 1103-1113.

[9]   Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *science*, *332*(6025), 60-65.

[10] Drosou, M., & Pitoura, E. (2015). Multiple radii disc diversity: Result diversification based on dissimilarity and coverage. *ACM Transactions on Database Systems (TODS)*, *40*(1), 1-43.

[11] Riaz, F., Alam, M., & Ali, A. (2017, August). Filtering the big data based on volume, variety and velocity by using Kalman filter recursive approach. In *2017 IEEE 3rd International Conference on Engineering Technologies and Social Sciences (ICETSS)* (pp. 1-6). IEEE.

[12] Graham-Rowe, D., Buxton, B., Hayward, V., Pearson, I., Karkkainen, L., Greiner, H., ...& Schillace, S. (2008). Big data: The next Google.

[13] Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in Healthcare: promise and potential. *Health information science and systems*, *2*(1), 3.

[14] Bahrami, M., & Singhal, M. (2015). The role of cloud computing architecture in big data. In *Information granularity, big data, and computational intelligence* (pp. 275-295). Springer, Cham.

[15] Yu, C., Lakshmanan, L., & Amer-Yahia, S. (2009, March). It takes variety to make a world: diversification in recommender systems. In *Proceedings of the 12th international conference on extending database technology: Advances in database technology* (pp. 368-378).

[16] Stephanidis, C., & Akoumianakis, D. (2002). Towards a design code of practice for universal access in health telematics. *Universal Access in the Information Society*, *1*(3), 223-226.

[17] Bidaut, L. M., & Scherrer, J. R. (1998). Telematics techniques for image based diagnosis, therapy planning and monitoring. *International journal of medical informatics*, *52*(1-3), 81-91.

[18] Liu, Z., Sun, P., & Chen, Y. (2009). Structured search result differentiation. *Proceedings of the VLDB Endowment*, *2*(1), 313-324.

[19] Grimson, J., Grimson, W., Berry, D., Stephens, G., Felton, E., Kalra, D., ... & Weier, O. W. (1998). A CORBA-based integration of distributed electronic healthcare records using the synapses approach. *IEEE Transactions on information technology in biomedicine*, *2*(3), 124-138.

[20] Pandey, R. V., & Schlötterer, C. (2013). DistMap: a toolkit for distributed short read mapping on a Hadoop cluster. *PloS one*, *8*(8).

[21] Langmead, B., Hansen, K. D., & Leek, J. T. (2010). Cloud-scale RNA-sequencing differential expression analysis with Myrna. *Genome biology*, *11*(8), R83.

[22] Bahrami, M., & Singhal, M. (2015). The role of cloud computing architecture in big data. In *Information granularity, big data, and computational intelligence* (pp. 275-295). Springer,Cha

[23] Feldman, B., Martin, E. M., & Skotnes, T. (2012). Big data in healthcare hype and hope. *Dr. Bonnie*, *360*, 122-125.

[24] Erkut, E., Ülküsal, Y., & Yenicerioğlu, O. (1994). A comparison of p-dispersion heuristics. *Computers & operations research*, *21*(10), 1103-1113.

[25] Vee, E., Srivastava, U., Shanmugasundaram, J., Bhat, P., & Yahia, S. A. (2008, April). Efficient computation of diverse query results. In *2008 IEEE 24th International Conference on Data Engineering* (pp. 228-236). IEEE.

[26] Agrawal, R., Gollapudi, S., Halverson, A., & Ieong, S. (2009, February). Diversifying search results. In *Proceedings of the second ACM international conference on web search and data mining* (pp. 5-14).

[27] Schatz, M. C. (2009). CloudBurst: highly sensitive read mapping with MapReduce. *Bioinformatics*, *25*(11), 1363-1369.

[28] Mohr, D. C., Burns, M. N., Schueller, S. M., Clarke, G., & Klinkman, M. (2013). Behavioral intervention technologies: evidence review and recommendations for future Research in mental health. *General hospital psychiatry*, *35*(4), 332-338.

[29] Zhu, X., Goldberg, A. B., Van Gael, J., & Andrzejewski, D. (2007, April). Improving diversity in ranking using absorbing random walks. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference* (pp. 97-104).

[30] Zhang, B., Li, H., Liu, Y., Ji, L., Xi, W., Fan, W., ... & Ma, W. Y. (2005, August). Improving web search results using affinity graph. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 504-511).

[31] Cox, Michael, and David Ellsworth. "Managing big data for scientific visualization." *ACM Siggraph*.Vol.97.1997.

[32] Christen, P. (2019). Data linkage: The big picture. *Harvard Data Science Review*.

[33] Hand, D., & Christen, P. (2018). A note on using the F-measure for evaluating record linkage algorithms. *Statistics and Computing*, *28*(3),539-547.

Authors' Profiles

Sidra Anwar was born in Sialkot, Pakistan in 1992. Now she is a PhD scholar in Memorial University of Newfoundland, Canada. Previously, she has been a lecturer in Computer Science at the Govt. College Women University, Sialkot, Pakistan and member of the ORIC Steering committee since 2015. She did her BS in Software Engineering from Fatima Jinnah Women University, the Mall, Rawalpindi, Pakistan and MS in Project Management from COMSATS Institute of Information Technology, Islamabad, Pakistan. She is also a Registered Engineer under Pakistan Engineering Council Act 1976 with Registration# COMP/14954. She has personally worked and supervised many projects regarding decision support systems serving multidisciplinary domains in IOT.

Iram Mehmood was born in Sialkot in 1998. She did her matriculation level in 2014 from, Govt. Girls Higher Secondary school Kharolian Sialkot in Science Subjects and her intermediate level (I.C.S) in 2016 from ILM collegeSialkot. Now she is doing her BS (Hons) in Computer Science (CS) from GC WOMEN UNIVERSITY, Sialkot. Currently she has been working on Research as final year project in Data Diversity inMedicalIoTtopic. Her main areas of research interest are Cloud architect, Computer architect,and Programming Languages.

Aneeza Dilawar was born in Sialkot in 1998. She did her matriculation level in 2014 from, Quaid Public High School Sialkot in Computer Arts and her intermediate level (I.C.S) in 2016 from ILM college Sialkot. Now she is doing her BS (Hons) in Computer Science (CS) from GC WOMEN UNIVERSITY, Sialkot. Currently, she has been working on Research as final year project in Data Diversity inMedicalIoTtopic. Her main areas of research interest are Cloud Computing, and Programming Languages.

Raja Manzar Abbas is a PhD scholor in University of Limerick, Ireland. Previously, he has been a lecturer in Software Engineering at Fatima Jinnah Woman University, Rawalpindi, Pakistan, and National University of Modern Languages (NUML), Islamabad. He did his BS in Computer Engineering from COMSATS Institute of Information Technology, Islamabad, Pakistan and MS in Software Engineering & Management from Gothenburg University, Gothenburg, Sweden. He has personally worked and supervised many projects regarding decision support systems and Digital Image Processing.

Isma Zulfiqar was born in Sialkot in 1998. She did her matriculation level in 2014 from Danish Public High School Sambrial, Sialkot in Science Subjects and her intermediate level (I.C.S) in 2016 from Superior College Sambrial, Sialkot. Now she is doing her BS (Hons) in Computer Science (CS) from GC WOMEN UNIVERSITY, Sialkot. Currently she has been working on Research as final year project in Data Diversity inMedicalIoT topic. Her main areas of research interest are Cloud Computing, primary dangerous computer viruses, Data Base and Programming Languages.