

HEALTH DATA SCIENCE: INSIGHT INTO THE RELEVANCE OF THE DOMAIN WITH EMERGING AND POSSIBLE ACADEMIC PROGRAMS- A KNOWLEDGE SURVEY

P.K. Paul*¹, A. Bhumali ², M.K. Ghose³, R. Rajesh⁴

*FBAS, Indian Institute of Engineering Science and Technology (IEST), Shibpur- An Institute of National Importance, Howrah, West Bengal, prancloud@outlook.com

²Vice Chancellor, Raiganj University, Raiganj, West Bengal, India

³Dean (Academics), Sikkim Manipal University, Gangtok, Sikkim, India

⁴Dean (Mechanical Science), NI University, Tamilnadu, India

Abstract

Keywords:

Health Data Science, Health Informatics, Big Data, Challenges, Public Health, Health Improvement, MSc-Health Data Science, Domain based Health Informatics, Big Data Management, Healthcare

Data Science and Big Data becomes valuable practicing domain having connections of scientific query. Though it also related with few other facets which includes statistical knowledge as well as substantive expertise moreover, computer programming is also important to note. It is a fact planner and policy makers today using several big data systems for greater productivity of the systems. Though it is fact that appropriate skills along with knowledge in Data Science in Health and Medical Science domain is still vacant. Substantive knowledge and also analytical skills is required for the scientists, economist with deep integration of quantitative methods. Health Data Science is a combination of Health Science and Data Science. Fundamental concepts and techniques of Big Data Analytics is also significant among the Health Data Scientist. Social, political, economic, legal, business, marketing etc professionals and skilled may join into Data Science profession apart from engineering sciences, though entering Health Sciences professionals will be an added advantage for the development of the domain. This is a conceptual paper and deal with several aspects of Health Data Science. The paper is specially highlighted the areas of skills, job potentials and current programs running in the domain of Health Data Science with possible programs in international context.

INTRODUCTION

Health Data Science is a domain of interdisciplinary in nature and combine with the Health Science and Data Science. It is a dealing of systematic as well as scientific processing of data, knowledge in the field of health and medical systems using intelligent systems [5], [9]. Medical Informatics and its advance form and areas in many cases considered as Health Data Science. Gradually the advancement of science and research has been revealed that the Health Informatics is much broader than Medical Informatics and in both the domains Health Data Science become common name in modern healthcare practice. It is a fact that Health Informatics is responsible for managing as well as dissemination of the information for sector. Collaboration and combination of the Information Science, IT, Computing, Medical Science, Management Science results Health Data Science as an intelligent domain [6], [8], [13]. Big Data integrates healthy and intelligent Medical Informatics practice and it helps in proper planning as well as implementation for first response and solutions. Many organizations in the world such as PwC, IBM, Accenture and Infosys are mapped this service in their catalogs to serve their clients. Internationally Business Analytics professionals are in high demand having skilled in Big Data Management and healthcare segment is most emerging. Courses and educational programs in Data Science seek diversity in several areas include management, science, technology, commerce etc. And program of Health Data Science may also offered in concentration of such domains [4], [7], [14]. The Health Big Data is normally offered to understand key concepts in addition to distinctions of the Health Data Science; that need to be synthesized for efficient and effective health data science integration. Role and aim of the health data scientist is emerging due to need of Medical Sciences. The Health Data Science professionals

required to learn how they fit into the broader healthcare scenery. Understanding of importance and values of patient-focused delivery along with outcomes are also important reason for Health Data Science introduction around the world.

OBJECTIVES WITH AGENDAS

As a conceptual study and knowledge survey this paper is deals with some of the aim and among them few important are include (but not limited to the)—

- To cram about Health Data Science including its principles, features, function at a quick look.
- To recognize and dig out the main advantages and reward of the Health Data Science and Healthcare in Big Data Management in current context.
- To learn about the higher education systems in the computing and informatics related areas around the world.
- To see the available Health Data Science programs in the universities and their utilizations in the corporate houses, industrial and sectors etc.
- To find out and describe main challenges, issues related to Health Data Science training and educational programs especially in the developing countries.

METHODOLOGIES

This is a conceptual framework based paper and ultimately responsible to know latest on Health Data Science in contemporary perspective. Mainly theoretical and conceptual research methodologies are used to make this policy paper. Collections of secondary data and also primary data have done from the review of literature. The websites of Health Informatics and Health Computation association have been utilized to learn latest on topic. Simple search strategy has used with the keyword MSc- Health Data Science and data/result of 15 pages considered as main source to learn the latest programs on Health Data Science. Current journals related to the Health Data Science, Big Data, Health Informatics, Computing, Tele-Medicine have been consulted to learn more about the domain.

HEALTH DATA SCIENCE: FUNDAMENTALS

Data Science is deals with the approaches to quantitative analysis of data. And for that several methods of statistical learning including the approach blending classical statistical methods have been deployed [1], [3], [15]. With recent advances and progress in the computational systems and ML the Health Data Science become more advance and progressing with proper practicum. Data Science is also called as Big Data Management program as large number of data basically analyzed and dealt by the Big Data Tools such as Hadoop. The Health Data Science is deals with analytical methods and practical tools and applications using example datasets. Data Science professionals needs practical gaining with latest methods of Statistics, Management and Database Systems. Health Data Science professionals have to deal the Data preparation, processing with databases (structured) formatted data and information. Importantly Health Data Science also handles unstructured and textual data which are valuable in most contexts in Data Science practicum. Big Data Management knowledgeable deemed and measured as an expert of Data Management and Mathematical approaches with understanding and ability of analyze data and similar content [2], [10], [16].

In Health Data Science use of few of its core and vital methods including solid basis are needed for more superior and more specialized study. Health Data Science is also develops the in-depth knowledge including the considerate and investigative skills with health data successfully to get better healthcare delivery and systems.

DATA SCIENCE AND EDUCATIONAL PROGRAMS EMPHASIZING HEALTH DATA SCIENCE

Data Science is creating several job opportunities around the world including in developing countries [11], [17], [20]. UK is the pioneer in Data Science applications and educated product generation. The territory is also responsible for strongest digital market creation and deployment in all sorts. Today every organization and institution basically rate by productively as a major competitive advantage. The development and digital economy is also depends on solid IT Applications. The development of United Kingdom is estimated larger in any other country as per major sources as

far as Digital Economy is concerned and here Data Science played a lead role. Thus Health Data Science also has a solid point to mention. There are many potential nomenclature in Health Data Science which include

- Health Data Management.
- Medical Big Data Systems.
- Health Data Analytics.
- Data and Health Intelligence.
- Health Data Science with Big Data.

‘Big Data’ predicted enormous financial value that may reach up to EUR250bn per year in each and every sector as far as McKinsey Global Institute is concerned. Department of Culture, Media and Sports, United Kingdom stated that “*The digital economy in UK is accounts for about £1 in each £10 that the UK market produces every year*”. Cross-research council ‘Digital Economy’ or ‘Connected Digital Economy Catapult’, etc are few programs and agenda that have solid interaction with Health Data Science [12], [21], [22]. Sustainable Development and growth outcomes are deemed as major for development of every kind. Healthy Digital Economy building is depends on technical skills including huge knowledge for managing including manipulating large datasets, interprets and represents them into knowledge. Hence knowledge of skills and technology and more clearly STEM are highly needed for creating a Data Scientist professional [18], [26]. McKinsey Global Institute depicted that ‘*Data Science is suffering with big data tools and technology*’. Health Data Science is utilizable in several settings including the industries, organizations, universities and so on [19], [23]. The Health Data Science is responsible for the following reason and include

- Health Data Science is responsible for gaining knowledge and an understanding of the healthcare system including the treatment of individuals (specific) to the wider sector and samples.
- Gain an thoughtful overview of health applications and information technology applications of the governance structures furthermore frameworks used when working with health data in the healthcare sector and area.
- To get experience as well as key technical skills along with software for working and analyzing and manipulating health data and records.
- To understand and get the breadth as well as depth of application methods and systems with the potential uses of health data.
- To understand as well as key concepts of distinctions with disciplines that need synthesized health related data.
- Appreciate and managing the responsibility of health data expert including the way of interring the technological integration in health segment. Here understanding of weight of the patient-focused delivery is also valuable.
- Health Data Science is needed to create in-depth knowledge as well as analytical and practical skills that are needed for managing health data effectively.
- Health Data Science is responsible for the development of a systematic and relevant knowledge, including theoretical frameworks. It also helps in shaping analytical skills to demonstrate the key challenges as well as issues [24], [25].
- Health Data Science is responsible for arising starting mixed data and volume and scale. Moreover it turns them for decision making in healthcare delivery including the research furthermore innovation and development.
- Health Data Science is also dedicated to applying the practical and skill based understanding to problems solving in healthcare and medical domain.
- Multi-disciplinary community serving and knowledge of using health data and record is also important agenda for developing Health Data Science as an interdisciplinary domain.
- Effectiveness and solid techniques, methods utilization in health challenges and issues are also vital for the introduction of Health Data Science as a program of study in emerging context.

- Health Data Science helps in understanding also the ability to the advancement of healthcare delivery of knowledge as well as research. It also helps in the practice and skills throughout, and here systematic, in-depth exploration in a medical field is important to keep in mind.

HEALTH DATA SCIENCE AND PROGRAM FEATURES—

Health Data Science is an interdisciplinary program and combines from the Information Science and Technology (i.e. *Computer Science, Information Studies, Information Science, Information Technology, Computing, Management Information Systems*), Health and Medical Sciences and Mathematical Sciences (i.e. *Mathematics-Pure and Applied, Statistical and Actuarial Sciences*) with adequate principles of Management Science [13]. Thus initially the Health Data Science deals the following domains and modules of health sciences include—

- Principle of Digital Biology
- Introduction to Health Informatics
- Fundamentals of Epidemiology
- Health Information Systems and Technologies
- Basics of Health Data Science
- Data and Decision Making Process and Management
- Fundamentals of Mathematics and Statistics for Health Data
- Biomedical Modeling for Health Data
- Foundation of Medical Diagnosis and Treatment

Health Data Science is focused with the computing and information technology branch and thus most of the programs deals with following modules as per major international universities curriculum.

- Data Base Management
- Web Systems and Health Data Dissemination
- Programming with Health Systems
- HCI and Usability Engineering in Healthcare

As Health Data Science is interdisciplinary in nature thus many universities have included following programs and modules into their curricula.

- Analytics and Statistical Programming
- Inferential Modeling of Health Outcomes
- Health Data Management and Standard
- High Performance Computing in Healthcare Industry
- Communication and Leadership in the Healthcare Industry
- Predictive Modeling and Machine Learning

HEALTH DATA SCIENCE AND JOB POTENTIALS

Data science and Big Data is growing rapidly around the world. Now, the domain is applicable in diverse fields such as medicine and public health, education and training, marketing and sales etc. The Health Data Science is needed to determine best design treatments as well as allocate resources depending upon health and related medical data. Educated in the field of Health Data Science and Health Big Data have the wider career opportunities including in the industry and academia [14], [16]. According to a study it is estimated that 52,000 data scientists required in the UK only. Educated and researcher in the fields may also go for research in several areas like Medicine and Health. One can also enter into Computing and Informatics related branches such as Computer Science, Information Studies, Information Science, Information Technology, Computing, Management Information Systems. The Health Data Science gives thorough quantitative training and techniques with the essential computing skills. Health Data Science is essential to manage as well as analyze health and medical science data to improve public health, medicine, and basic biology applications too.

HEALTH DATA SCIENCE AND SOME RUNNING UNIVERSITIES

The Educational programs around the world changing into its new shape and thus several interdisciplinary programs are coming in many new universities. The universities of UK and US based universities are providing courses in Health Data Science. Though as per our search strategy adopted Universities in United Kingdom are pioneer. The Table: 1 is depicted herewith with listing of courses.

Table: 1 Some programs in Health Data Science according to the knowledge Survey

| Universities | Running Programs |
|--------------------------------------|---|
| Swansea University | MSc Health-Data Science |
| The University of Manchester, London | MSc Health-Data Science |
| University College London, London | MSc-Data Science for Research in Health and Biomedicine |
| Saint Louis University, Spain | MS- Health Data Science |
| Harvard University, US | MSc- Health Data Science |
| Lancaster University, UK | MSc-Data Science (Health) |

HEALTH BIG DATA: POSSIBLE PROGRAMS

There are many programs available in Health Data Science and Health Data Management. Most of them are mainly with the tag of MSc/MS- Health Data Science. Availability of BSc/BS- Health Data Science or in similar nomenclature becomes tough. Thus there are opportunities to offer possible programs in other areas as a specialization. The Table: 2, Table: 3, Table: 4 and Table: 5.

In Science Stream

In Science, including Basic Science and Applied Science, the Health Data Science may positively possible to start. The Table: 2 herewith depicted. Programs on Computing have also depicted.

Table: 2 Some possible Health Data Science programs in the context of Science domain/s.

| Possible Bachelor Degrees | Possible Masters Degrees |
|---|--|
| BS/BSc- Health Data Science & Informatics | MS/MSc- IT & Health Data Science |
| BS/BSc-Medical Informatics with Big Data | MS/MSc -Medical Information Sciences & Big data |
| BSc- Information Technology (Health Data Analytics) | MSc- Information Technology (Big Data) |
| BSc-Computer Science (Health Informatics and Analytics) | MSc-Computer Science (Data Analytics & Management) |
| BSc-Computer Application (Medical Big Data) | MSc-Information Science (Health Data Analytics) |
| | MSc-Computer Application (Data Analytics & Telemedicine) |

In Engineering Stream

In the Engineering and Technological stream i.e. BTech (Bachelor of Technology), BE (Bachelor of Engineering) etc the Health Big Data may included in several platform. ME/MTech degrees may also offer with Health Data Science specialization. Learn Table: 3 to more about the proposed programs in the areas.

Table: 3 Health Data Science domains Vis-à-Vis possible Engineering Degrees

| Possible Bachelor Degrees | Possible Masters Degrees |
|---|--|
| BTech/BE- Health Informatics & Analytics | MTech/ME- IT & Health Data Sciences |
| BTech/BE -Medical Information Sciences with Big Data | MTech/ME –Health Data Science & Cheomo-informatics |
| BTech/BE - Information Technology/ Computer Engineering (Health Big Data) | MTech/ME - Information Technology (Health Analytics & Cloud) |
| | MTech/ME -Computer Science (Medical Data Management & Tele-systems) |

In Management

Medical Analytics may start easily as specialization in Health Information Science. There are healthy potential offer Management Science based Health Analytics programs (as depicted in the Table: 4).

Table: 4 Possible Management Degrees in Health Science flavor

| Possible Bachelor Degrees | Possible Masters Degrees |
|---|---|
| BBA (Health Analytics) | MBA (Health Data Analytics & Informatics) |
| BBA (Health Information Management with Big Data) | MBA (Health Data Science) |
| BBM (Health & Data Analytics) | MBM (Hospital Informatics with Big Data) |
| B.Com (Health Analytics & Management) | M.Com (Health Analytics & Management) |

In Computer Application—

Application of the computing and information technology in widen areas are the core task of Computer Application. Post Graduate Diploma, Bachelors, Masters, Doctoral levels etc are offered in Computer Application domain. Health Data Science may easily start in Computer Application program. The possible programs are (In Table: 5)—

Table: 5 Possible Health Data Science in Computer Application flavor

| Possible Bachelor Degrees | Possible Masters Degrees |
|--|---|
| BCA (Health Analytics) | MCA (Health Information Systems and Analytics) |
| BCA (Medical Informatics and Big Data) | MCA (Medical Informatics and Big Data) |
| BCA (Health Analytics with Telemedicine) | MCA (Health & Medical Data Science) |
| | MCA (Medical Information Science with Big Data) |

FINDINGS

- Interdisciplinary domain and combines with the Health Data Science even the philosophy of arts, science and technology etc also have included in the domain.
- Health Data Science is the broader domain than Health Big Data. Apart from these it is also known as Health-Big Data Management, Health Data Analytics, Health and Medical Data Analytics.
- Health Informatics is the core branch of Health Data Science and the domain is increasing rapidly around the world.
- In the Health Data Science several tools of IT also become integrated part and some of them are include Cloud Computing, Human Computer Interaction.
- Health Data Science is mainly available as MSc degree in United Kingdom. Health Informatics focused programs are also offered.

- Health Data Science is also offered with Medical Informatics nomenclature/s. The program popular in Health Science, Computer Science and Management Science Schools/ Departments.
- In the developing and undeveloped nations like India, Bangladesh, Malaysia, South Africa the Health Data Science program are still absent as full-fledged program.

SUGGESTIONS

- Association and cooperation for the universities, governments, related scientific foundations etc are urgently required for implementing Health Data Science in reality.
- For the promotion and development Health Data Science and allied domains and fields the small and higher level of education need to start.
- The specializations in the field may also be started in the related departments such as computing, health science, management as depicted in Table: 2 onwards.
- The running Health Informatics projects need to carry in the Health Data Science domain for its proper development and for that contemporary training; seminars, workshop etc need to organize.
- Health Informatics is an important branch and thus it is very urgent that universities and research centre should gear up for collaborative and interdisciplinary research in several level.

CONCLUSIONS

Health Data Science is a valuable domain, tools and in interdisciplinary health and medical science. Healthcare science and medical science need to gear-up for advancement of research for a solid and sustainable development. The Health Informatics as well as Health Data Science implementation is increasing in academic as well as industrial segment. Empowerment as well as steps needs to take properly for creating a modern advancement in healthcare systems. Proper planning as well as policies implementation are urgently required for real development of management, statistical, information technology, information systems applications in the healthcare system. Developing nations also need to join hands with the developed countries to provide sophisticated and intelligent health systems.

REFERENCES

1. Aarts, J., Peel, V., & Wright, G. (1998). *Organizational issues in health informatics: a model approach. International journal of medical informatics*, 52(1), 235-242.
2. Friede, A., Blum, H. L., & McDonald, M. (1995). *Public health informatics: how information-age technology can strengthen public health. Annual review of public health*, 16(1), 239-252.
3. Haux, R. (1997). *Aims and tasks of medical informatics. International journal of medical informatics*, 44(1), 9-20.
4. Hersh, W. R., Junium, K., Mailhot, M., & Tidmarsh, P. (2001). *Implementation and evaluation of a medical informatics distance education program. Journal of the American Medical Informatics Association*, 8(6), 570-584.
5. Hook, S. A. (2003). *Teaching health informatics: designing a course for a new graduate informatics program. Journal of the Medical Library Association*, 91(4), 490.
6. Kun, L. G. (2001). *Telehealth and the global health network in the 21 st century. From homecare to public health informatics. Computer methods and programs in biomedicine*, 64(3), 155-167.
7. Martin, S.B. (1998). *Information technology, employment, and the information sector: Trends in information employment 1970-1995. Journal of the American Society for Information Science*, 49(12), 1053-1069.
8. Michael Buckland and Ziming liu (1995). *History of information science. Annual Review of Information Science and Technology vol. 30: 385-416.*
9. Paul, P.K., (2012b) "Medical 2.0: The tool for proper Information Transfer Cycle (ITC) among the Medical Professionals and Physician with special reference to its difference with Health 2.0" *International Journal of FEE* 4 (2), 183-190
10. Paul, P.K., Ashok Kumar, Dipak Chaterjee (2012) " Health Informatics and its Practice: Emerging Domain of Information Science-Indian Scenario" in *Current Trends in Biotechnology and Chemical Research*, 2 (2), 83-87

11. Paul, P.K., D Chatterjee, M Ghosh (2012) "Medical Information Science: Emerging Domain of Information Science and Technology (IST) for sophisticated Health & Medical Infrastructure Building — An Overview" in *International Scientific Journal of Sports Science*, 1 (2), -104
12. Paul, P.K., D Chatterjee, M Ghosh (2012b) "Neural Networks: Emphasizing its Application in the World of Health and Medical Sciences" *Journal of Advances in Medicine*, 1 (2), 17-23
13. Paul, P.K. (2013) "CFTRI- Knowledge Hub: The Wonderful Information and Knowledge Resource Center of Food and Nutrition Science in India" in *Abhinav National Journal of Science and Technology*, 2 (3), 37-42
14. Paul, P.K., M Ghosh (2013) "Web 2.0 and its Possible, Emerging and Future Domain Related to Health and Medical Sciences" *Journal of Information Technology and Engineering*, 4 (2), 51-55
15. Paul, P.K., M Ghosh (2013b) "Cloud Computing and its possible utilization in Health and Hospital Administration" *Journal of Business Management [JBM]- An International Journal*, 5 (2), 147-152
16. Paul, P.K., J. Ganguly, M Ghosh (2013) "Chemo Informatics: the Healthy Results of Information Science and Chemical Science Integration" *International Journal of Computer Science Systems Engineering and Information Technology*, 6 (2), 92-96
17. Paul, P.K., (2013) "Food and Nutrition Information Systems in India: Challenges and Issues" *International Journal of Computer Science and Knowledge Engineering*, 7 (2), 134-138
18. Paul, P.K., J. Ganguly, M Ghosh, (2013) "Medical Information Science: Overview and a model curriculum of MSc-Information Science [Medical Information Science]" Submitted for *Current Trends in Biotechnology and Chemical Research*, 3 (1), 50-54
19. Paul, P.K., (2013) "Web 2.0 and Social Networking in Pharmacy and Allied Medical Domain: An overview" *International Journal of Marketing Theory*, 3 (2), 137-143
20. Paul, P.K., (2013) "Web 2.0: emphasizing Pharmacy 2.0/ Pharmaceutical 2.0-Promising Interdisciplinary Information Science Field" *Scholars Academic Journal of Pharmacy (SAJP)*, 2(6):476-480
21. Saracevic, T. (1979b). *An essay on the past and future of information science education. II. Unresolved problems of 'externalities' of education* *Information Processing & Management*, 15(4), 291-301.
22. Vakkari, S.P. (1996). *Library and information science: Content and scope*. In J. Olaisen, E. Munch-Petersen, & P. Wilson (Eds.), *Information science: From development of the discipline to social interaction*. Oslo, Norway: Scandinavian University Press.
23. Vickery, B.C., & Vickery, A. (1987). *Information science in theory and practice*. London: Butterworths.
24. Wersig, G., & Neveling, U. (1975). *The phenomena of interest to information science*. *Information Scientist*, 9, 127-140.
25. Yasnoff, W. A., O'Carroll, P. W., Koo, D., Linkins, R. W., & Kilbourne, E. M. (2000). *Public health informatics: improving and transforming public health in the information age*. *Journal of Public Health Management and Practice*, 6(6), 67-75.
26. https://en.wikipedia.org/wiki/Health_informatics (Accessed on 05-08-2016)