

## Data Science in Healthcare

Snober Jon<sup>1</sup>, Shafqat Manzoor<sup>1</sup>, Beenish Bashir<sup>1</sup>, Monisa Nazir<sup>2</sup>

<sup>1</sup>Student, Computer Science and Engineering,

<sup>2</sup>Assistant Professor, Department of Computer Science Engineering,

<sup>1,2</sup>S.S.M College of Engineering & Technology, Kashmir, India

### ABSTRACT

The main aim of this paper is to provide a deep analysis on the research field of healthcare data analytics., as well as highlighting some of guidelines and gaps in previous studies. This study has focused on searching relevant papers about healthcare analytics by searching in seven popular databases such as google scholar and springer using specific keywords, in order to understand the healthcare topic and conduct our literature review. The paper has listed some data analytics tools and techniques that have been used to improve healthcare performance in many areas such as: medical operations, reports, decision making, and prediction and prevention system. Moreover, the systematic review has showed an interesting demographic of fields of publication, research approaches, as well as outlined some of the possible reasons and issues associated with healthcare data analytics, based on geographical distribution theme.

**KEYWORDS:** Healthcare, Data Analytics, Clinics, Systematic Review, Tools and Techniques

### 1. INTRODUCTION

The rapid development in technology facilitated the healthcare system to generate huge amounts of information related to the patient followed by their records pertaining to disease, treatment history, test reports, etc.

The concern towards healthcare is increasing day by day with the rapid increase in population. Traditionally data is available in the form of hard copies, but due to technology development and in the era of digital world everything is collected and stored in the digitized form. It can be predicted that in the future there will be significant growth in the health data. The other organization like insurance companies, government organizations are using this data to provide health benefits to the patients. Analysis of disease patterns, tracking the disease can help us to find a cure for the disease, and dealing with disease eruption can eradicate the chances of its growth. Such analysis can enhance the level of public health and awareness and rapid actions to control the diseases. Development of required vaccines by the medical researchers. Converting large amounts of

**How to cite this paper:** Snober Jon | Shafqat Manzoor | Beenish Bashir | Monisa Nazir "Data Science in Healthcare" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-6 | Issue-1, December 2021, pp.617-621, URL: www.ijtsrd.com/papers/ijtsrd47870.pdf



Copyright © 2021 by author(s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



health information into predictive models to recognize the needs, the services to be provided, predict, and prevent a health disaster for the benefit of people. Improvement in operational functionality can also add light to the healthcare system. For instance, by investigating patient history, frequency of patient visits, and analyzing staff efficiency, healthcare facilities can optimally allocate healthcare staff to a particular shift without having to overstaff or under staff. Predictive analytics is vital to achieving the goal of providing better care and cutting down on healthcare costs simultaneously.

With the use of disease predictive modeling, virtual assistance can also be provided to patients. Based on the confidence rate, with very basic information input about symptoms patients can get insights about the various possible diseases. Drug discovery is very time and cost consuming. Mutation profiles and patient metadata can be used by various pharmaceutical companies for the discovery of new drugs using Machine learning Algorithms. One of the critical elements in the examination and figuring out the

treatment strategy is medical image analysis. Mammography, X-ray, MRI, and others provide vital insights into the patient’s disease. Accuracy of the images and its interpretation needs to be top-notch in medical Image analysis. Supervised machine learning algorithms can help in medical image analysis by improving its Image resolution, modality difference, and the dimension of images.

**2. BACKGROUND**

Data is an imperative aspect of every industry and organization. Due to the internet, around 2TB of data is being generated by the users every day. This data includes several types of forms that may be either structured or unstructured.

In the earlier, the Doctors were not able to monitor the patients’ health condition in real time. As disease increases, the cost to cure it also increases, so to cure the extreme diseases was very expensive.

Nowadays, with the help of Data Science and Machine Learning applications, Doctors can get real-time information about the patients’ health condition through their wearable devices. Based on this information, the Doctors can send the junior doctors, nurses, or assistants to the patient’s home.

Not only the Doctors can provide immediate assistance and treatment to the patients, but they can also install several required equipment and devices for the patient’s diagnosis. These devices can collect heart rate, body temperature, blood pressure, etc., information through updates and notification of mobile applications on the top of Data Science.

**3. METHODS**

The objective of this paper was to conduct a review, which encourages professionals, doctors, medical staff and patients to adopt and utilize technologies in order to assist healthcare analytics and improve decision making process in our everyday life.

Our method has followed three steps: 1) searching for initial and related studies, 2) Relevance appraisal and evaluation, and finally extracting data. The next sections will explain these steps briefly.

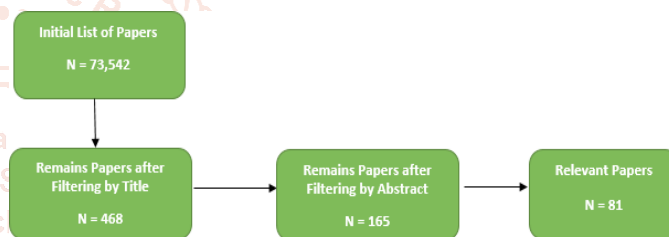
Searching for initial and related studies: the first step in order to find the articles was to specify and identify main keywords (Dieste et al. 2009). A survey was conducted to study relevant papers published since 2010 in the information system field in general and healthcare analytics and medical decision support system in specific. This study has found that most

relevant keywords to “healthcare analytics” and “data mining” used with technology to support medical information systems.

The following searching phrases were used and structured in searching for relevant papers in many different databases – i.e. the relevant and related papers should contain in its titles, keywords, abstract or full text the word “healthcare” along with any of “analytics”, “metrics”, “data mining”, “big data” or “decision making”. see table 1

Group 1	Group 2
E-Healthcare	Analytics
Medical Practice	Metrics
Health	Decision Making
Clinical	Prediction
Hospitals	Big Data
Care systems	Data Mining
Wellness programs	Business Intelligence

**Table: Main keyword**



**Fig.1 Search Process**

As a result, searching techniques relied on choosing any key word from the (group A) and linked it with any word of (group B) to form searching keywords statement, such as (“E-Healthcare” OR “Medical Practice” OR “Health” OR “Clinical” OR “Hospitals” OR “Care Systems” OR “Wellness Programs”) AND “Analytics” OR “Metrics” OR “Decision Making” OR “Prediction” OR “Big Data” OR “Data Mining” OR “Business Intelligence”.

Once the keywords were identified, 7 online databases were searched to find the initial list of the studies. In the search, titles, keywords, abstract and full text were considered and the search was limited to studies published since 2010, inclusive.

The databases were searched over multiple subjects and returned total of 73,542 articles (see Figure 1). This study found some of the papers are indexed by multiple databases. As it shown in Table 2, total number of the papers after deducting the repeated papers was 81.

Name of Database	Initial list of papers	Papers Filtered By Title		Papers Filtered By Abstract		Papers Filtered By Text	
		Number of found articles	Number of repeated articles	Number of found articles	Number of repeated articles	Number of found articles	Number of repeated articles
Google Scholar	37,000	72	11	47	11	25	9
IEEEExplor	6,791	103	5	39	5	20	5
ACM Digital Library	2041	49	6	21	4	7	2
ProQuest	16,139	133	8	44	5	31	4
Scopus (Elsevier)	1,380	104	3	29	3	14	3
Springer	8568	43	9	15	4	6	1
Wiley InterScience	1623	11	5	7	5	3	1
<b>Total</b>	<b>73,542</b>	<b>468</b>		<b>165</b>		<b>81</b>	

Table 2: Initial list of studies

## 4. RESULTS

### 4.1. Fields of Publication

This systematic review has found Information Systems with 43 papers and Healthcare with 31 papers as most active communities in the research related to the healthcare analytics topic, however 7 papers were published related to the healthcare analytics in Computer Science. Figure 3 shows that authors focused on information systems and healthcare fields more than computer science.

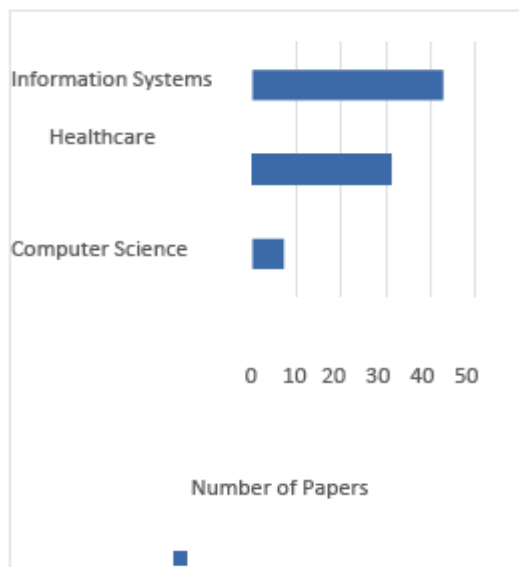
As it shown in Figure 2, academic papers related to the healthcare analytics and decision making were mostly published in information systems and healthcare for the reason that most studies recently have focused on improving healthcare analytics using data mining and business intelligence techniques, however a few were published in the field of computer science. This could be because of computer scientist were dealing with old traditional methods and trying to solve general issues using these methods rather than suggesting new techniques due to the evolution of technology in these days, as well as most of articles were trying to use old methods and involving only doctors and professionals in healthcare analytics process in order to improve clinical and hospitals performance as an interest topic in computer science, paying no attention for the importance to involve patients in that.

### 4.2. Application Areas

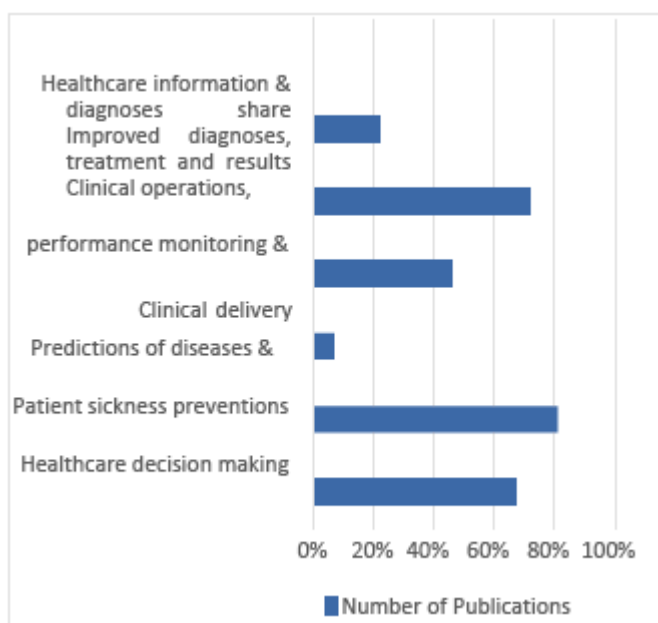
This paper found that healthcare analytics papers has focused on six main areas: (a) healthcare decision making (b) predictions of diseases & patient sickness preventions (c) clinical delivery (d) clinical operations, performance monitoring & reporting (e) Improved diagnoses, treatment and results and finally (f) Healthcare information exchange. However, there was some studies focused on financial and supply chain management, but due to the small number of publications in these areas, we removed them from this Figure.

As it shown in Figure 3 most attention was paid to improve healthcare analytics performance and results, therefore the Figure below illustrates that most publications focused healthcare prediction and preventions, decision making process and healthcare treatment and monitoring.

For many researchers, the main factors of reaching high level of healthcare analytics were in simplifying unstructured clinical records, as well as capturing patient's behavior and encourage individuals to educate themselves, as well as keep following ups, by adopting technologies and internet services and applications, for example social networking sites & social media which may allow them to keep updated and connected with other patients worldwide, sharing their health information and supporting each other in order to improve healthcare analytics and reduce costs.



**Figure 2: Fields of Publications**



**Figure 3: Application Areas**

## 5. ADVANTAGES

- Support clinical treatment decisions from physicians and other health professionals.
- Improve the accuracy and speed of identifying patients at highest risk of disease.
- Provide greater detail in the EHRs of individual patients.
- Make the provision of healthcare more efficient, which reduces costs.
- Promote preventive measures by giving patients greater insight into their health and treatment goals.
- Integrate data from consumer fitness devices and other patient-provided sources of health data.
- Deliver real-time alerts to healthcare providers by analyzing health data at the collection point.

## 6. FUTURE SCOPE

There have been many improvements done in the healthcare sector, but still, some more applications and improvements are required in the future like:

- Digitalization
- Technological Inclusion
- Reduced Cost of Treatment
- Need to be able to handle huge amount of patient’s information

Data science tools and technologies are working for these requirements and have made many improvements as well. Data science is doing wonders in many real-life areas and contributing a lot. There will be much assistance available for doctors and patients through this revolution of data science in the future.

## 7. CONCLUSION

Finally, we conclude that data science has many applications in healthcare. The healthcare industry is heavily dependent on Data Science for its improvement. Additionally, with the development and advancement of medical image analysis, it will be possible for physicians to find microscopic tumors that are in fact difficult to find manually. Therefore, data science has revolutionized healthcare and the large-scale medical industry.

## 8. REFERENCES

- [1] Laney D. 3D data management: controlling data volume, velocity, and variety, Application delivery strategies. Stamford: META Group Inc; 2001. Google Scholar
- [2] Mauro AD, Greco M, Grimaldi M. A formal definition of big data based on its essential features. *Libr Rev.* 2016;65(3):122–35. Article Google Scholar
- [3] Gubbi J, et al. Internet of Things (IoT): a vision, architectural elements, and future directions. *Future Gener Comput Syst.* 2013;29(7):1645–60. Article Google Scholar
- [4] Doyle-Lindrud S. The evolution of the electronic health record. *Clin J Oncol Nurs.* 2015;19(2):153–4. Article Google Scholar
- [5] Gillum RF. From papyrus to the electronic tablet: a brief history of the clinical medical record with lessons for the digital Age. *Am J Med.* 2013;126(10):853–7. Article Google Scholar
- [6] Reiser SJ. The clinical record in medicine part 1: learning from cases\*. *Ann Intern Med.* 1991;114(10):902–7. Article Google Scholar

- [7] Reisman M. EHRs: the challenge of making electronic data usable and interoperable. *Pharm Ther.* 2017;42(9):572–5. Google Scholar
- [8] Murphy G, Hanken MA, Waters K. *Electronic health records: changing the vision.* Philadelphia: Saunders W B Co; 1999. p. 627. Google Scholar
- [9] Shameer K, et al. Translational bioinformatics in the era of real-time biomedical, health care and wellness data streams. *Brief Bioinform.* 2017;18(1):105–24. Article Google Scholar
- [10] Service, R.F. The race for the \$1000 genome. *Science.* 2006;311(5767):1544–6. Article Google Scholar

