

# Connected Patients in Light of Big Data

THE JOURNAL 2022

Diogo Neves, Henrique Martins  
Linking Patients With Data: Brain-Computer  
Interfaces and Healthcare Innovation

Eugene Fidelis Soh et al.  
Flipping Healthcare Through a Population  
Health Stack

Srdjan Babic et al.  
Data Science in Modern Healthcare

Francisco Maestre et al.  
DICOM Metadata - A Useful Resource for  
Big Data Analytics

Jörg Schwarz  
How to Utilise the Massive Amount of  
Health Data Collected by Consumers to  
Improve Health Outcomes

Betsabe Melcon et al.  
In Search of Gold in Health Data





# Data Science in Modern Healthcare

**Srdjan Babic** | Institute for Cardiovascular Diseases Dedinje | University of Belgrade Faculty of Medicine | Belgrade, Serbia

**Masa Petrovic** | Institute for Cardiovascular Diseases Dedinje | University of Belgrade Faculty of Medicine | Belgrade, Serbia

**Branko Lozuk** | Institute for Cardiovascular Diseases Dedinje | University of Banja Luka Faculty of Medicine | Belgrade, Serbia

**Milovan Bojic** | Institute for Cardiovascular Diseases Dedinje | Belgrade, Serbia

Data science has brought forth many exciting advancements in healthcare; however, it is important to recognise what data science is and how we can continue to use it to positively advance healthcare.

## Key Points

- Data science is an intersection of computer science, mathematics, and domain expertise.
- Artificial intelligence encompasses machine learning and deep learning which are the building blocks of data science.
- Data science has several applications in modern healthcare which can aid in decision making, lead to earlier diagnosis, and allow us to better practice evidence-based medicine as physicians.
- When sharing data it is important to know the FAIR guiding principles and the differences in databases that are open access, selected access and closed access.

## Introduction

Modern society has evolved over time with the help of computers and technology to gather data and be used to open avenues for further research. In order to understand how data is being used in healthcare to improve outcomes and patient quality of life every day, it is important to dig deep to understand the roots of data science and data management.

## What is Data Science?

In order to understand what exactly data science is, it is important to take a step back and understand what a database is. A database is an organised collection of data that is stored and accessible electronically. They can range from being small to large, requiring different storage methods. A small database can be stored on a file, while larger databases can be stored on computer clusters, cloud storage, and servers. Without a database, data science would not be possible.

Data science is a field of study that is focused on dealing with vast volumes of data using modern techniques and tools

to find patterns, obtain meaningful information, and use this information to aid in decision making. Data science can also be looked at as an intersection of computer science, mathematics, and domain expertise. Within this intersection lays software, data analysis, and machine learning. Data science can also be considered the powerhouse of the evolution of artificial intelligence which can be split into three categories: artificial intelligence, machine learning, and deep learning. Data science encapsulates machine learning and deep learning.

## Machine Learning vs Deep Learning vs Artificial Intelligence

As we step forward into a society where there is an increasing use of machine learning, deep learning, and artificial intelligence, the terms are often wrongly used interchangeably. Hence, it is important to make a clear distinction between the three terms and encourage the correct usage.

Data science encompasses deep learning and machine learning. Beyond that, artificial intelligence encompasses all



three (data science, deep learning, and machine learning). Given we have defined what data science is, we will start from small and level up to the big picture of artificial intelligence.

First, deep learning is a subset of machine learning where vast amounts of data stored in databases is continually analysed and based on this information a layered structure of algorithms, artificial neural network, is used for the system to “learn.” Next, machine learning is where the performance of those algorithms can improve as they are exposed

shared, and closed. These types of databases mainly vary in the way the data is shared and who has access to the database.

On one end of the spectrum is open data where data collected is available to everyone to use, view, and even publish based on this data. For data to be truly considered open, it is pivotal that there are no restrictions from copyright, patents, or other mechanisms of control. Some benefits of open data is that its use and re-use can provide data to researchers and

---

## Data science can be considered the powerhouse of the evolution of artificial intelligence which can be split into three categories: artificial intelligence, machine learning, and deep learning

---

to more data over time. Finally, artificial intelligence is where a programme can use the information from deep learning and machine learning to act upon the data, adapt, sense patterns, and aid in decision making.

Over the past few years, data science has an increasing presence in the healthcare field with the introduction of smart-watches with ECG monitoring capabilities, computer-aided diagnostics, and several other technologies. These technologies in healthcare can help us reach a diagnosis at earlier stages, point out abnormalities on a medical image that would otherwise be missed by the human eye, save time, and even aid in decision making.

With the proper use of artificial intelligence in healthcare, we can improve safety, quality, and outcomes of future patients. As physicians, we can practice evidence-based medicine and use artificial intelligence as an extra tool to help guide decisions and even use the data constantly being collected to update guidelines and improve outcomes.

Within data science it is important to recognise the importance of sharing data which can lead to further advancement of artificial intelligence. In healthcare environments where there is a lot of confidential data, it is important to recognise some key guiding principles that should be followed for sharing, and how data should be stored and shared.

### Data Science and Healthcare: To Share or Not to Share?

In data science there are common guiding principles that should be followed (FAIR). FAIR is **f**indable, **a**ccessible, **i**nteroperable, **r**eusable. Furthermore, it is pivotal to consider the privacy and discoverability of data. In data collection, it is important to consider how data is shared. Particularly in healthcare management, this is a widely debated topic.

Databases can be classified into three different types: open,

other data scientists who traditionally would not have access to this type of data. Open data can positively contribute to research in that this data can be viewed by many researchers that can offer varied perspective on the data and draw the most out of it. While there are many advantages to open data, there are a lot of concerns that surround this type of managed data when it comes to humans, and particularly in healthcare management.

In healthcare, it is pivotal to pay attention to confidentiality and privacy and usage can be limited. In addition, in regard to human data, open access databases can be commonly misused and even be interpreted by individuals that do not have proper knowledge in the field.

On the other far end of the spectrum is closed data. This is where data can only be viewed by those within a particular organisation where the data is being collected. This also means that the data is not shared beyond the organisation.

In between open and closed data is shared data. This is where data is closed with the exception that it can be shared only within a certain group of people for a specific purpose.

Within healthcare, it is critical for centres to have their own data collection and a closed data approach. In the case of a multicentre collaboration for a research project, data can potentially be considered shared given that certain precautions are taken to ensure that privacy is maintained for sensitive data (data that can be used to identify an individual, species, object, process or location that introduces a risk of discrimination, harm or unwanted attention).

When working with sensitive data, before sharing the data, it is pivotal to hide any parts of it (identifiable) to make it non-identifiable. This is to ensure that the data cannot be misused, and that patient data remains confidential.

In regard to human research, the data collected can be categorised in three forms: identifiable, re-identifiable, and



non-identifiable. Identifiable is where based on the data, it is very clear who the patient is. Re-identifiable is where patient data is de-identified or anonymised to a certain degree; however, it is still possible to identify an individual if

understand how this occurs. The advancement of technology along with the creation of databases have given rise to a new field of study, data science, which has entered into various sectors, particularly healthcare. As data science continues to

---

## In data science, there are common guiding principles that should be followed - findable, accessible, interoperable, reusable

---

you have access to all of the data. For example, in healthcare management, this would be the difference between having the patient's name versus a number that their name and other sensitive information is linked to. Lastly, there is nonidentifiable data where data is de-identified/anonymised to the degree that by using the data the specific individual cannot be identified. When considering the sharing of data, especially with human data it is important that dependent on the degree that the data is shared the data is properly anonymised.

### Conclusion

In today's modern world where data is constantly being collected and analysed using algorithms, it is pivotal to

evolve within healthcare to aid in decision making and help improve safety and outcomes for patients, it is important to recognise the keen differences between artificial intelligence, machine learning, and deep learning as well as how data should be stored and shared. While the future of data science in healthcare may be very exciting, it is crucial for us to not lose sight of the importance of keeping patient data confidential, refraining from misusing them, and understanding the limitations of such technologies.

### Conflict of Interest

None. ■



**HealthManagement.org**

*Promoting Management and Leadership*