IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 8 Issue 12, December 2021
ISSN (Online) 2348 – 7968 | Impact Factor (2020) – 6.72
www.ijiset.com

The Application of Big Data Analytics in Medicine and Healthcare for Clinical Decision Support Systems

Imane ELFALOUSSI[1]

[1] Institute of graduate studies/Department of Computer Engineering/ Information Technologies, Altınbaş University, Turkey.
imane.elfaloussi@gmail.com

Abstract

Nowadays, Big data has provided a considerable revolution in the healthcare field; it helped generate a relationship between science life and healthcare, leading to connecting doctors, pharmaceuticals, and patients. This revolution will disrupt out health systems and will transform the methods of care which will lead ultimately to influence our public health policy. Particularly, Big Data in healthcare represents the huge immense and complex data that are related with the healthcare system and which are not easy to be managed or analysed with traditional methods.

The analyses of the data collected from patients permit to integrate the healthcare field with many other scientific areas, such as bioinformatics, medical informatics and health informatics. And also, can allow moving from a curative care system to a preventive care system by providing recurring systems, detecting the undesirable effects of drugs or the misuse of them and helping as a support clinical decision system.

Keywords:

Big data, Clinical decision support systems, CDSS, Exploratory Data Analysis, Healthcare, Pandas, Python.

1. Introduction

Big Data in the field of health references to the patient's records, collected from different resources to benefit from many advantages like: identification of diseases risk factors, helping in diagnosis, monitoring the effectiveness of treatments, medicine, epidemiology, and pharmacovigilance. In fact, there are many ways to collect medical data, either via health professionals (Patient's medical records during hospitalization or doctor's visit, or through clinical trials or DNA analysis), or via individuals themselves (Individuals can produce their own data through connected objects such as watches, mobile applications etc.). These data generated in various ways produce a strong correlation between the different actors of health system. We can even make patient remote monitoring which was impossible before.

The increasing use of data in medicine is gradually leading to switch from making diagnoses based on symptoms to making prognoses based on the patient's history. In this new context a number of scientific, ethical and legal questions arise. It is then a question of evaluating and managing a risk rather than pathology according to medical, financial, ethical and social criteria, which can lead to preventive actions and/or prescriptions that impact the patient's lifestyle. Nevertheless, Big Data in health raises many challenges and constitute many ethical questions.

1.1 Big Data analytics in Medicine

Big Data technology provides many facilities in many fields including healthcare, the collected data of patients helps doctors to make preventions and correct decisions [1]. It is easy to collect Data in Healthcare context, either from tests results, pharmaceutical data, clinical trials, electronic health records, mobile apps or sensors [2].

There are four Vs primary attribute that identify the Big Data; Volume, Velocity, Variety and Veracity. Volume refers to the size, Velocity to the speed, Variety to data collected from different sources, while Veracity refers to the quality; a data with a high veracity is a data that has many valuable records to be analysed. A low veracity data is a data that contains meaningless or valueless records [3].

-Clinical Data characteristics:

A patient's HER (Electronic Health records) has a significant role in health research since it provides an important number of clinical data. Over the last few years, we are witnessing a huge growth of data, and it is due to the smart wearable devices that can collect people's health data every day and at any moment (like blood pressure, heart rate etc.) [4].
-Clinical Data Types: During a treatment process of a patient, a huge number of information is collected and generated, which means that the clinical data is much diversified that it can be divided into three categories: Clinical notes, medical images, and all others.
Clinical notes:

In EHRs, text is considered as one of the most important categories, it provides rich information related to the patient's health status such as test results, diagnoses, treatment and drugs. Besides that, other facultative data are also recorded that may help in clinical decision support systems and prevent from disease or misuse of treatment like family medical history, allergies, etc.
Medical images:

Medical images is the process that helps imaging the interior of a patient's body for clinical analysis, it is obtained commonly from X-rays, Magnetic Resonance Imaging (MRI), microscopy image, Optical Coherence Tomography (OCT), or Positron Emission Tomography (PET).[5]
Other types: Clinical data doesn't come only in medical notes or images form, there are many other types, such as information about payment or insurance, physiological measurements result and demographic results (gender, age, location, and marital status), laboratory's tests results, which are all represented with numeric values [6].

1.2 Introduction to CDSS

An adult can make many decisions per day, some of them may not really matter, but in healthcare field, it is not the same, as one right decision can save a life while the wrong decision can cause critical consequences.
Clinical Data are expanding more and more, and to achieve full benefits of electronic health records, a good quality CDSS is highly recommended. CDSS or Clinical Decision Support System is considered as a program based on computer which helps Healthcare personals improving patients care by analysing data and making right clinical decisions. It is an adaptation of DSS (Decision Support System) that is used normally in business management. CDSS use knowledge management based on patient's data to provide care plan and assistance recommendations.
To make the best decisions for patients, health personals must control diverse details and factors under pressure in a short time. Now, fortunately, they can do most part of this job with computers, precisely with Clinical Decision Support Systems. The first CDSS in the world was built at Stanford University in 1970s, named MYCIN, but due to its slow performance, plus ethical and legal issues, it wasn't put in practice.
Today, Healthcare uses CDSS for many different tasks such as drugs control, ordering tests and generating alerts. There are different CDSS; standalone CDSS that can focus only on one problem for example sending reminding alerts, or a CDSS that can assure a large variety of processes [7].

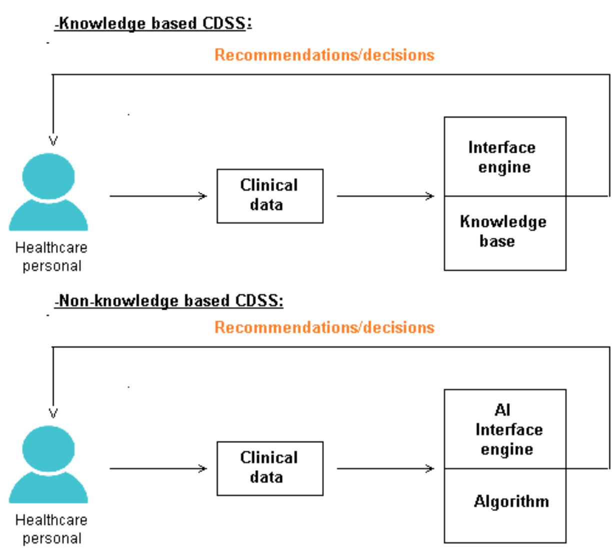


Fig. 1. Knowledge-based and non-knowledge based CDSS

According to the computer science methodology, there are two main groups of Clinical Decision Support Systems; Knowledge-based system and non-knowledge-based system [8].

-Knowledge-Based CDSS: contains 3 main parts, a knowledge base, inference rules and mechanism of communication. It is based on IF-THEN statements (IF the medicine A is taken, and medicine B is taken too at the same time, THEN the user will get an alert).

-Non-Knowledge-Based CDSS: it is a system without a knowledge base, based on artificial intelligence, which mean that the computer machine finds patterns in clinical data by learning from past experiences [9].

2. Methods and algorithms

2.1 Application of Big data analysis in healthcare.

Generous amount of data is generated in systems of healthcare (patient's demographic, drugs, patient's medical history, insurance information, etc.) which are stored digitally. The use of this big data analysis with efficacious techniques impacts positively the healthcare systems and help in making right decisions. Nazari et al. (2021). [10] inspected the framework of the six building blocks that the health system is based on. Which are: service delivery, Health workforce, Health information systems, Access to essential medicine, financing and leadership/governance. Nazari et al. revealed that considering this framework and its dimensions is necessary to provide good healthcare services. Minu et al. (2017). [11] showed that the analysis of data in real-time is very important in prediction of emergency cases and clarified that is necessary to use Hadoop cluster approaches either the data will be useless.

2.2 Privacy issues of big data in healthcare.

As the number and the volume of medical data in increasing rapidly, the number of security and privacy issues is increasing as well. Leyngar et al. (2018) [12] said that the big data do a great role in healthcare since it can provide opportunities for development of medical care, but the privacy of patients, doctors and healthcare's personals is more important and concerned. Daniels et al. (2018) [13] inspected the heterogeneous data in healthcare and ontology, and conclude that the privacy issue in medical data is quite serious.

2.3 Solution for Big Data privacy issues in healthcare.

Recently, experts and researchers investigated many methods and technologies to solve the problem of privacy in big data. Fritchman et al. (2018) [14] discussed that in healthcare field, strict agreements need to be applied for sharing data, also strong encryption is recommended to realize the data protection. Gruscka et al. (2018) [15] said that personal data are not only the personal identifiers (Full name, ID number) but also include indirect identifiers and information such as phone number, address, photos etc. And focused on the GDPR (General Data Protection Regulation) which considered as the most important development created to protect data for all individuals in the EU in this century. Bhuiyan et al. (2018). [16] suggested a system which is based on block chain technology, a technology that record data in a way that is impossible to be changed or hacked, which provide high protection to the medical data.

2.4 The implementation of a CDSS.

The first step for developing a CDSS begins with data since data come in different forms. The categorical data or structured data (numeric data) are easy to work with. The data used for CDSS includes medical registries which are developed for research purposes and clinical trials that have an important role for predefined research questions.

Before using data to build a model, it is necessary to pre-process the data; it is a step where variables are defined from raw data that a model can use. It is also important to manage the valuable and relevant variables from data by the experts of healthcare. After the extraction of meaningful data, the next step to build a model of CDSS is to use an algorithm. Usually, depending on the task's complexity, the development of the CDSS models contains two phases (Model training and model validation), the first phase is for developing a model that fits data, while the second phase is for testing if the model is correct and suitable for the tasks wanted.[17]

2.5 Functionalities of CDSS.

The use of big data analysis for Clinical Decision Support Systems provides many functionalities and benefits to healthcare system.[18]

-Patient safety: CDSS is capable to reduce the risk of medications errors, researches show that up to %65 of patients have been exposed at least one time to wrong drugs combinations [19]. CDSS successfully reduce the prescriptions or the dosages errors [20].
-Clinical management: CDSS shows its ability to organize and assist patient on treatment protocols [21]. And can also send alert if patient didn't follow management plans.
-Cost containment: CDSS is considered as a practical and worthwhile for healthcare systems [22], since it provides services like reducing test duplication, and decreasing timespan of inpatients.
-Administrative functions: CDSS can help physicians by suggesting diagnostics codes using designed algorithms.
-Diagnostics support: known also as DDSS (Diagnostic Decision Support System) is a system that present a consultation by using data given to return a number of diagnoses possible [23]. This system, unlike the CDSS couldn't much influence in healthcare system due to physician's negative discrimination [24].
Patient decision support: CDSS come up with a dashboard where patients can communicate with doctors, get informed about treatments or medications. Additionally, CDSS can be used for actionable insights by collecting data from application or wearable devices used by patients [25].

The real aim of Clinical Decision Support System is to make the right decisions by involving a strong relation between physicians, patients and scientific evidences. Recently, it is been witnessed a progressive change promoting shared decisions made by physicians or health professionals and patients. The structure of shared decisions permits healthcare personals to build a relation with patients to choose the right tests and treatment based on patient's choices and the scientific proofs [26].

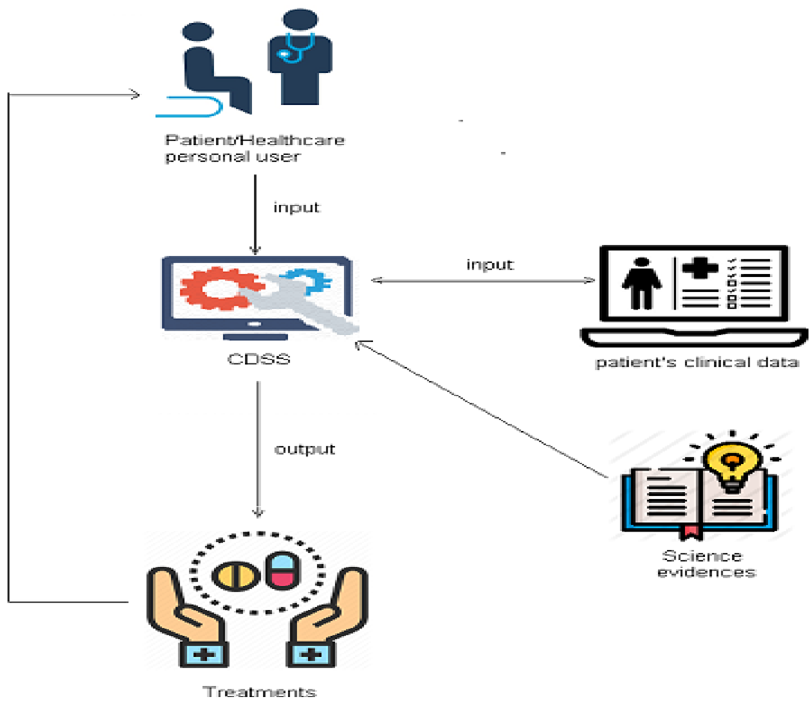


Fig. 2. Clinical decision support system Diagram.

The figure above presents a diagram that describes how CDSS works in healthcare field. The first step starts with the healthcare personal/ patient who provide information needed, then the CDSS analyses the data. The CDSS according to the patient clinical data (personal information, medical history, test, etc.), plus the domain knowledge (science evidences, researches) output the suitable decisions and treatments.

2.6 CDSS in different medical specialties.

Nephrology is one of the most important medical specialties, but the use of Big data and CDSS is still limited. Since CDSS is deemed as an important tool for decision-making based on evidences, algorithms and models are necessary to get good results [27]. A modern CDSS based on web was built by Samal et al. (2017) [28] that can evaluate the risk of kidney failure using data collected from CKD (Chronic kidney disease) patients. But this model still limited in many ways; it is asynchronous, since it cannot support a real time decision-making and requires patient's information at least the night before the visit.

Oncology is also one of the foremost medical fields that can benefit from the application of big data in Medicine system and the implementation of CDSSs to fulfill good results. One of the most known CDSS is IBM Watson For Oncology (WFO), it is based on artificial intelligence that can help cancer patients by providing therapeutic options. NCCN (National Comprehensive Cancer Network) is integrated by WFO to provide guidelines of treatment and therapeutic options related to cancer and supply links of evidences published by scientific researches [29].

There are some programs that support people attaint by different cancers. CancerLinQ is a program that collects data from different sources then organizes the data to be used by healthcare personals that are responsible of cancer patients' care [30]. OncoAnalytics supply physicians with necessary information about care tuitions and drug doses [31]. And lastly, Tempus program, which have a huge clinical data related to precision medicine that healthcare professionals can use to facilitate their decision-making [32].

CDSS programs are not dedicated only to clinicians, but also to patients, like iCanDecide, a program for people who have breast cancer where they can navigate, learn about treatment and record their preferences [33]. Decision Board is also one of the programs made for cancer patients, it is a platform where doctors explain about treatment options and patient can participate in the decision-making.

3.Exploratory Data Analysis

The statistician John W. Tukey [1977] defined the exploratory data analysis as an approach that analyzes and examines data sets to discover their main characteristics and to see what the data can communicate us. EDA focuses on 4 keys; the measurement of central tendency, the measurement of spread, the shape of distribution and the existence of outliers.
The EDA approach makes the manipulation of data sources easier for data scientists and helps to get valid results that can be applicable to any desired goal. It helps to well understand patterns, identify errors and make interesting relations between variables.

1. The application of EDA

Exploratory Data analysis can be used to detect anomalies and outliers, to test assumptions, and to gain accurate perceptions of data.

3.1 EDA in Python

Exploratory Data Analysis in Python is considered as the first and main step in the process of data analysis. It was developed by the statistician John Tukey in the 1970s. Python is used for exploratory data analysis because it is easy and simple to learn, it is an open-source language with high data handling capacity, can be run on many platforms, with numerous and rich libraries [34].
Jupyter notebook:
Jupyter notebook is a server-client application that uses web browser to allow running notebook documents. It can be either executed on local desktop without need to internet. Or can be used via internet after being installed on a remote server.

Pandas:
Pandas is a Python library, it is a powerful open-source data analysis tool, it is fast, flexible and most used for data science/ data analysis and machine learning tasks. Panda is really a game changer when it comes to manipulating data. It helps on cleaning, storing and visualizing data easily.

3. Application of EDA on a DATA SET

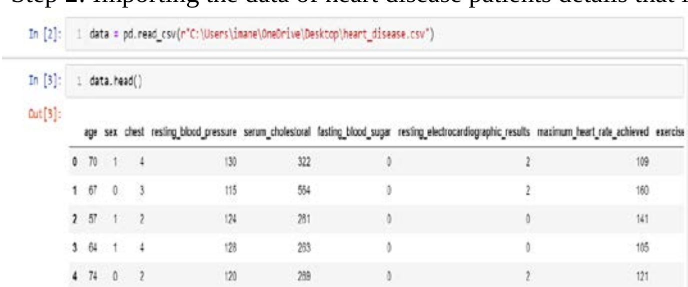
In this chapter, we are going to apply the Exploratory Data analysis on a Data set imported from Kaggle website using Python, Jupyter notebook and Pandas library.

Used data set: heart disease patients' details.

Attributes:
-Age: represents the age of a patient (from 1 to 100)
-Sex: represents the gender (0: Female, 1: male)
-Chest: represents the chest pain of a patient
-Resting_blood_pressure: represents the blood pressure of a patient while resting. (Normal: below 120/80 mm or elevated: less than 80 mm Hg or from 120 to 129 mm Hg)
-Serum_cholestoral: represents the measurement of certain elements in the blood (high- and low-density lipoprotein cholesterol (HDL & LDL))
-Fasting_blood_sugar: represents the concentration of glucose in the blood (Normal: between 70mg/dl = 3.9 mmol/L and 100mg/dl=5.6 mmol/L, while fasting: between 100mg/dl = 5.6 mmol/L to 125mg/dl=6.9 mmol/L)
-Resting_electrocardiographic_results: represents the results of a test that take 5 to 10 mins. No movement is allowed during the test to not impact the results of the electrical impulses generated by the heart.
-Maximum_heart_rate_achieved: representing the maximum number of times that a heart should beat per minute while exercising, which can be obtained from subtracting age from 220.
-Exercice_during_angina: physical activity triggers the angina. While walking, exercising or climbing stairs, the heart demands more blood.
-Old_peak: Exercise relative to rest.
-Number_of_major_vessels: In heart there are 5 major blood vessels (aorta, the superior vena cava, the inferior vena cava, the pulmonary artery, and the pulmonary vein).
-Thal: Thalassemia is a disorder in blood that causes a deficiency in body's hemoglobin (Hemoglobin carries oxygen to the organs by enabling red blood cells)
-Result: represents if the patient has a heart disease or no.
It is time to explore the data set of heart disease patient's details using Python and its different libraries.
-Step1: Importing Pandas library, numpy package, seaborn (a Python library for visualizing data in an informative statistical graphics), and matplotlib.pyplot(to provide a MATLAB-like way of plotting).

Fig. 3. Importing Pandas library and necessaire packages

-Step 2: Importing the data of heart disease patients details that is stored in a CSV file. And visualizing the first 5 rows.



	age	sex	chest	resting_blood_pressure	serum_cholestoral	fasting_blood_sugar	resting_electrocardiographic_results	maximum_heart_rate_achieved	exercise
0	70	1	4	130	322	0	2	109	
1	67	0	3	115	564	0	2	160	
2	57	1	2	124	261	0	0	141	
3	64	1	4	128	263	0	0	105	
4	74	0	2	120	269	0	2	121	

Fig. 4. Importing data file and showing the 5 first rows of dataset.

-Step 3: starting the exploration and visualization of data using different functions.

IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 8 Issue 12, December 2021
ISSN (Online) 2348 – 7968 | Impact Factor (2020) – 6.72
www.ijiset.com

In [5]:
```
1 data.describe()
```

Out[5]:

	age	sex	chest	resting_blood_pressure	serum_cholestoral	fasting_blood_sugar	resting_electrocardiographic_results	maximum_heart
count	270.000000	270.000000	270.000000	270.000000	270.000000	270.000000	270.000000	
mean	54.433333	0.677778	3.174074	131.344444	249.659259	0.148148	1.022222	
std	9.109067	0.468195	0.950090	17.861608	51.686237	0.355906	0.997891	
min	29.000000	0.000000	1.000000	94.000000	126.000000	0.000000	0.000000	
25%	48.000000	0.000000	3.000000	120.000000	213.000000	0.000000	0.000000	
50%	55.000000	1.000000	3.000000	130.000000	245.000000	0.000000	2.000000	
75%	61.000000	1.000000	4.000000	140.000000	280.000000	0.000000	2.000000	
max	77.000000	1.000000	4.000000	200.000000	564.000000	1.000000	2.000000	

Fig. 5. A summary describes of data frame.

In [7]:
```
1 data.columns
```

Out[7]:
```
Index(['age', 'sex', 'chest', 'resting_blood_pressure', 'serum_cholestoral',
       'fasting_blood_sugar', 'resting_electrocardiographic_results',
       'maximum_heart_rate_achieved', 'exercise_induced_angina', 'oldpeak',
       'slope', 'number_of_major_vessels', 'thal', 'result'],
      dtype='object')
```

Fig. 6. the columns of the dataset.

In [11]:
```
1 data.dtypes
```

Out[11]:
```
age                                       int64
sex                                       int64
chest                                     int64
resting_blood_pressure                    int64
serum_cholestoral                         int64
fasting_blood_sugar                       int64
resting_electrocardiographic_results      int64
maximum_heart_rate_achieved               int64
exercise_induced_angina                   int64
oldpeak                                 float64
slope                                     int64
number_of_major_vessels                   int64
thal                                      int64
result                                    int64
dtype: object
```

Fig. 7. Types of the attributes

In [35]:
```
sns.regplot(x="serum_cholestoral", y="maximum_heart_rate_achieved", data=data)
plt.ylim(0,)
```
Out[35]: (0.0, 208.55)

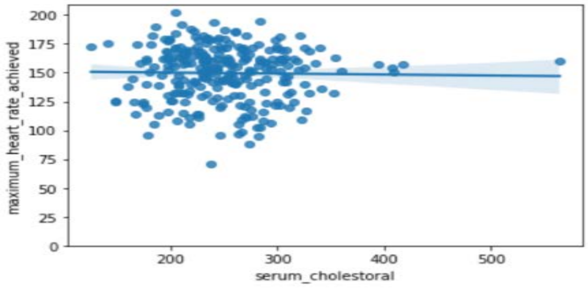


Fig. 8. Reg plot and scatter plot between maximum_heart_rate_achieved and serum_cholestoral

In [67]:
```
sns.heatmap(corr, annot=True)
```
Out[67]: <AxesSubplot:>

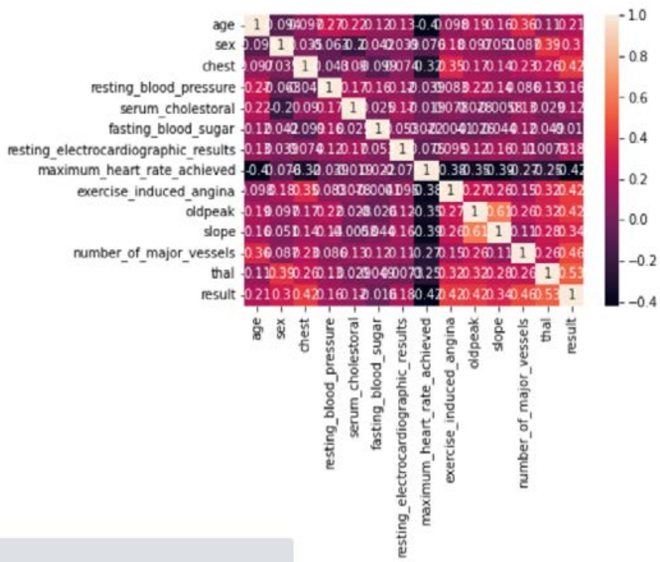


Fig. 9. Heatmap of the data set.

In [68]:
```
sns.countplot(x='sex', hue='fasting_blood_sugar', data=data)
```
Out[68]: <AxesSubplot:xlabel='sex', ylabel='count'>

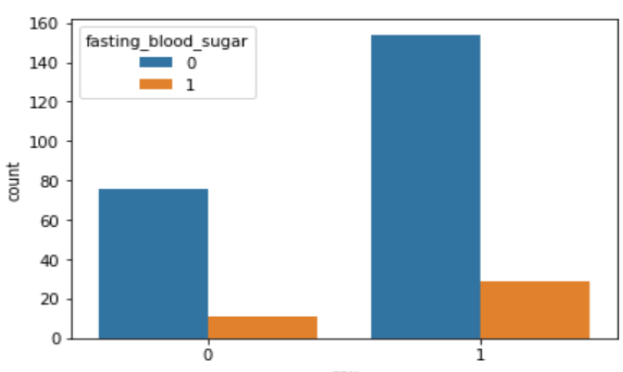


Fig. 10. Count plot between sex and resting_blood_sugar

```
In [25]: plt.figure(figsize=(15,5))
         sns.countplot(x=data['age'], hue=data['result'])
         plt.xticks(rotation=90);
```

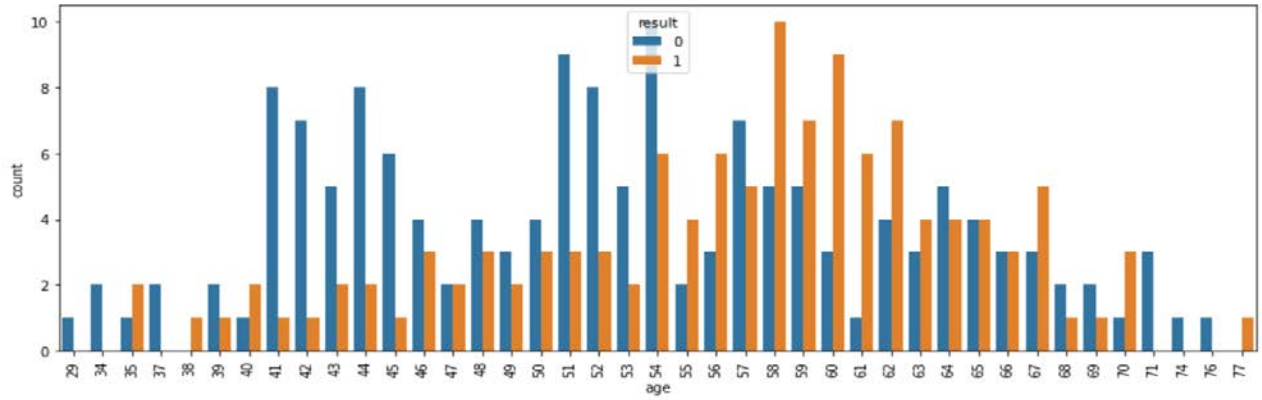


Fig. 11. The occurrence of heart disease patients according to age

```
In [37]: sns.relplot(x='result', y='age', hue='sex', data=data)
Out[37]: <seaborn.axisgrid.FacetGrid at 0x1e9a342ab20>
```

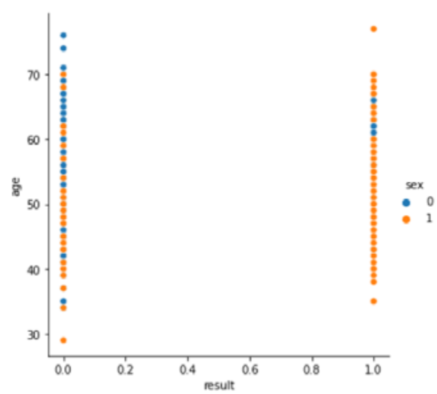


Fig. 12. The occurrence of heart disease patients according to age and sex(gender)

```
In [27]: sns.histplot(data['age'])
Out[27]: <AxesSubplot:xlabel='age', ylabel='Count'>
```

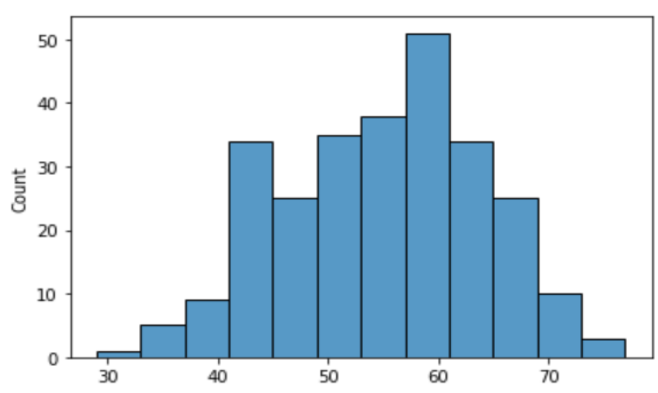


Fig .13. Histogram of patient's age

The figures [4,5,6,7,8,9,10,11,12,13,14 &15] show how using Exploratory Data Analysis can help in exploring data related to healthcare (heart disease patients in our case) and well understand it to make important benefits and good decisions. Various functions used while the exploration of data;
-Correlation: Define the pairwise of data columns, Figure-9.
-Regplot: Plot data and help emphasizing patterns in a dataset while exploring data, Figure-10.
-Heatmap: Shows higher values with dark color shades, and lower values with lighter shades, Figure-11.
-Countplot: Uses bars to show the observation counts in each categorical bin, Figure-12, Figure-13.
-Relplot: Uses semantic mapping of subsets to draw relational plots between two variables, Figure-14.
-Histplot: Based on matplotlib in Python, it represents the distribution and variation in data, Figure-15.

4. Conclusion

Big Data analytics is used almost in all sectors, and healthcare is considered as the most benefitted. collecting data in healthcare can be time consuming with traditional methods, but thanks to technology, today it becomes easy to collect and store data, and also use it to improve clinicians and healthcare system.

The researches made in this paper indicated that the application of big data analytics in medicine improved the quality of this field and enhanced the decision-making process. Additionally, the use of Clinical Decision support systems is considered as a good weapon that contributed to develop healthcare system against diseases. Admittedly, there still the problem related to the privacy that is not completely solved, but the collaboration of big data analysis, CDSSs, image processing and deep learning can help in diagnosis of many vital diseases such as chronic kidney diseases, viral bacterial infections [35], heart attacks, and cancers [36,37,38,39,40]. Additionally, the changes in human's life standards effect its metabolism and can provoke some rare diseases, allergies and disorders, which mean there will be always a need to technology in healthcare systems, it is a confrontation against the illnesses that may threaten human's life [41,42].

The research approached also the application of Exploratory Data analysis in healthcare which is really important, since its purpose is to better understand data, verify obvious errors, provide relations between variable and detect if there are any anomalous events.

Acknowledgments

The authors of this paper would like to show appreciation to all the other authors of the same topic for providing information and codes of their proposed works.

References

[1] Lv, Zhihan, and Liang Qiao. "Analysis of healthcare big data." Future Generation Computer Systems 109 (2020): 103-110.
[2] Saunders, Gabrielle H., Jeppe H. Christensen, Johanna Gutenberg, Niels H. Pontoppidan, Andrew Smith, George Spanoudakis, and Doris-Eva Bamiou, "Application of big data to support evidence-based public health policy decision-making for hearing." Ear and hearing 4, no. 5 (2020): 1057.
[3] S. Kumar and M. Singh, "Big data analytics for healthcare industry: impact, applications, and tools," in *Big Data Mining and Analytics*, vol. 2, no. 1, pp. 48-57, March 2019, doi: 10.26599/BDMA.2018.9020031.
[4] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. Med Image Anal. 2017 Dec;42:60-88. doi: 10.1016/j.media.2017.07.005. Epub 2017 Jul 26. PMID: 28778026.
[5] Akhtar, Usman, et al. "The Impact of Big Data In Healthcare Analytics." *2020 International Conference on Information Networking (ICOIN)*. IEEE, 2020.
[6] Yu, Ying, Min Li, Liangliang Liu, Yaohang Li, and Jianxin Wang. "Clinical big data and deep learning: Applications, challenges, and future outlooks." *Big Data Mining and Analytics* 2, no. 4 (2019): 288-305.
[7] Sharma, M., G. Singh, and R. Singh. "Clinical decision support system query optimizer using hybrid Firefly and controlled Genetic Algorithm." *Journal of King Saud University-Computer and Information Sciences* (2018)
[8] Wasylewicz, A. T. M., and A. M. J. W. Scheepers-Hoeks. "Clinical decision support systems." Fundamentals of clinical data science (2019): 153-169

[9] Sobowale, A. A., O. M. Olaniyan, O. Adetan, J. B. Oladosu, S. O. Olabiyisi, and E. O. Omidiora. "DEVELOPMENT OF FUZZY RULES FOR CDSS BASED NEONATAL MONITORING SYSTEM."
[10] Nazari, Elham, et al. "Application of Big Data analysis in healthcare based on 'The 6 building blocks of health systems' Framework: A survey." *Dokkyo Journal of Medical Sciences* 48 (2021): 01.
[11] Minu, M. S., Ishan Meena, R. Pratyush, and Vijayditya Sarker. "Healthcare Analysis Using Hadoop Framework." *leadership* 7: 11.
[12] A. Iyengar, A. Kundu, G. Pallis, Healthcare informatics and privacy, IEEE Internet Comput. 22 (2) (2018) 29–31
[13] M. Daniels, J. Rose, C. Farkas, Protecting patients' data: An efficient method for health data privacy, in: Proceedings of the 13th International Conference on Availability, Reliability and Security, ACM, 2018, p. 9
[14] K. Fritchman, K. Saminathan, R. Dowsley, et al., Privacy-preserving scoring of tree ensembles: A novel framework for AI in healthcare, in: 2018 IEEE International Conference on Big Data, 2018, pp. 2413–2422.
[15] Gruschka, N., Mavroeidis, V., Vishi, K., & Jensen, M. (2018). *Privacy Issues and Data Protection in Big Data: A Case Study Analysis under GDPR. 2018 IEEE International Conference on Big Data (Big Data).* doi:10.1109/bigdata.2018.8622621
[16] M.Z.A. Bhuiyan, A. Zaman, T. Wang, et al., Blockchain and big data to transform the healthcare, in: Proceedings of the International Conference on Data Processing and Applications, ACM, 2018, pp. 62–68.
[17] El Naqa, Issam, et al. "Prospects and challenges for clinical decision support in the era of big data." *JCO clinical cancer informatics* 2 (2018): 1-12.
[18] Sutton, Reed T., et al. "An overview of clinical decision support systems: benefits, risks, and strategies for success." *NPJ digital medicine* 3.1 (2020): 1-10.
[19] Vonbach, P., Dubied, A., Krähenbühl, S. & Beer, J. H. Prevalence of drug-drug interactions at hospital entry and during hospital stay of patients in internal medicine. Eur. J. Intern. Med. 19, 413–420 (2008)
[20] Jia, P., Zhang, L., Chen, J., Zhao, P. & Zhang, M. The effects of clinical decision support systems on medication safety: an overview. PLoS ONE 11, 1–17 (2016).
[21] Lipton, J. A. et al. Impact of an alerting clinical decision support system for glucose control on protocol compliance and glycemic control in the intensive cardiac care unit. Diabetes Technol. Ther. 13, 343–349 (2011).
[22] Calloway, S., Akilo, H. & Bierman, K. Impact of a clinical decision support system on pharmacy clinical interventions, documentation efforts, and costs. Hosp. Pharm. 48, 744–752 (2013)
[23] Berner E. Clinical Decision Support Systems: Theory and Practice 3rd edn. https:// doi.org/10.1007/978-0-387-38319-4 (2016).
[24] Segal, M. M. et al. Experience with integrating diagnostic decision support software with electronic health records: benefits versus risks of information sharing. EGEMs Gener. Evid. Methods Improv. Patient Outcomes 5, 23 (2017).
[25] Benhamou, P. Y. Improving diabetes management with electronic health records and patients' health records. Diabetes Metab. 37(Suppl. 4), S53–S56 (2011).
[26] Pereira, A. M., Jácome, C., Amaral, R., Jacinto, T., & Fonseca, J. A. (2019). *Real-Time Clinical Decision Support at the Point of Care. Implementing Precision Medicine in Best Practices of Chronic Airway Diseases, 125–133.* doi:10.1016/b978-0-12-813471-9.00022-0
[27] Yang, Chao, et al. "Big data in nephrology: Are we ready for the change?." *Nephrology* 24.11 (2019): 1097-1102.
[28] Samal L, D'Amore JD, Bates DW, Wright A. Implementation of a scalable, web-based, automated clinical decision support risk-prediction tool for chronic kidney disease using C-CDA and application programming interfaces. J. Am. Med. Inform. Assoc. 2017; 24: 1111–5.
[29] Suwanvecho, Suthida, Harit Suwanrusme, Tanawat Jirakulaporn, Surasit Issarachai, Nimit Taechakraichana, Palita Lungchukiet, Wimolrat Decha et al. "Comparison of an oncology clinical decision-support system's recommendations with actual treatment decisions." *Journal of the American Medical Informatics Association* 28, no. 4 (2021): 832-838.
[30] Rubinstein, Samuel M., and Jeremy L. Warner. "CancerLinQ: origins, implementation, and future directions." *JCO clinical cancer informatics* 2 (2018): 1-7.
[31] Dutton, Gail. "Making Every Cancer Care Outcome Count: Flatiron Health Aggregates Cancer Treatment Data from Clinical Trials and the "Real World"." *Genetic Engineering & Biotechnology News* 37.4 (2017): 6-7.
[32] Kwon, Diana. "Tempus and Mayo Clinic Partner for Data-Driven Cancer Care." *Clinical OMICs* 4.1 (2017): 14-14.
[33] Hawley, Sarah T., et al. "Improving breast cancer surgical treatment decision making: the iCanDecide randomized clinical trial." *Journal of Clinical Oncology* 36.7 (2018): 659.
[34] Sahoo, K., Samal, A. K., Pramanik, J., & Pani, S. K. (2019). Exploratory data analysis using Python. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *8*(12), 2019.
[35] S.C. Groft, M.P. de la Paz, Preparing for the future of rare diseases, in *Rare Diseases Epidemiology: Update and Overview* (Springer, Cham, 2017), pp. 641–648

[36] Manogaran, Gunasekaran, et al. "Machine learning based big data processing framework for cancer diagnosis using hidden Markov model and GM clustering." *Wireless personal communications* 102.3 (2018): 2099-2116.
[37] Ismail, Ahmed, Samir Abdlerazek, and I. M. El-Henawy. "Big data analytics in heart diseases prediction." *Journal of Theoretical and Applied Information Technology* 98.11 (2020).
[38] Teramoto, Atsushi, et al. "Decision support system for lung cancer using PET/CT and microscopic images." *Deep Learning in Medical Image Analysis* (2020): 73-94.
[39] Al-Saadi MJ, Ilyas M. Identity Management Approach in Internet of Things (IoT). In2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) 2020 Oct 22 (pp. 1-6). IEEE.
[40] Alkurdi DA, Ilyas M, Jamil A. Cancer detection using deep learning techniques. Evolutionary Intelligence. 2021 Jul 2:1-9.
[41] Ilyas M, Ucan ON, Bayat O, Nasir AA, Imran MA, Alomainy A, Abbasi QH. Evaluation of ultra -wideband in radio channel and its effects on system performance. Transactions on Emerging Telecommunications Technologies. 2019 Jan;30(1):e3530.
[42] M. Ilyas, O. N. Ucan, O. Bayat, X. Yang and Q. H. Abbasi, "Mathematical Modeling of Ultra-Wideband in Vivo Radio Channel," in *IEEE Access*, vol. 6, pp. 20848-20854, 2018, doi: 10.1109/ACCESS.2018.2823741.

Author

Imane ELFALOUSSI received the B.S degree in Mathematics and Informatics from Cadi Ayyad University, Safi, Morocco, in 2017. And a M.S degree student in Information technologies at Altinbaş University, Istanbul, Turkey (2020-2022).