Computer Science Review 37 (2020) 100280

Contents lists available at ScienceDirect

Computer Science Review

journal homepage: www.elsevier.com/locate/cosrev

Review article

When machine learning meets medical world: Current status and future challenges

Abir Smiti

LARODEC, Université de Tunis, Institut Supérieur de Gestion de Tunis, 41 Avenue de la liberté, cité Bouchoucha, 2000 Le Bardo, Tunisia

ARTICLE INFO

Article history:
Received 2 April 2020
Received in revised form 26 May 2020
Accepted 8 June 2020
Available online xxxx

Keywords:
Artificial intelligence
Machine learning
Medical field
Healthcare

ABSTRACT

Imagine the enormous amounts of data that can be generated in the medical field. Each patient has his own medical record which contains valuable information like patient allergy, chronic diseases and vaccinations. Healthcare can profit from this data when it is properly analyzed. The more data gathered, the more complicated data analytics become, therefore, machine learning can be a very useful solution not only to facilitate analysis but also to save time.

In this paper basic concepts of the medical field and machine learning will be described. We will show how data analytic can help in the healthcare process. Finally, we will present some challenges that must be carefully studied in order to obtain effective solutions in medical diagnosis.

Contents

E-mail address: smiti.abir@gmail.com.

https://doi.org/10.1016/j.cosrev.2020.100280
1574-0137/© 2020 Elsevier Inc. All rights reserved.

1. Introduction

One of the main differences between humans and computers resides on the fact that humans learn from past experience whereas computers needs to be told what to do, needs to be programmed so they follow instructions. Nowadays with Machine learning, we can get computers to learn from experiences too. For computers experiences refers to data. And as we know the world is filled with data, a lot of data: pictures, music, words, spreadsheets, videos etc. And it does not looks like it is going to slow down anytime soon. Machine learning brings the promise of deriving meaning from all that data. There are a lot of data in the world today generated not only by people, but also by computers, phones, and other devices. This will only continue to grow in the years to come. We see machine learning all around us in the products we use today. Today machine learning's immediate applications are yet quite wide-ranging including image recognition, text and speech systems too, fraud detection as well as recommendation systems. These powerful capabilities can be applied to a wide range of fields, from diabetic retinopathy and skin cancer detection to retail and of course, transportation in the form of self-parking and self-driving vehicles. Machine learning offers effective methods, techniques and tools, to help dealing with a big number of problems in various domains. Especially, clustering one of the famous techniques of machine learning, which has been extensively used for knowledge discovery from massive data.

In this paper basic concepts of the medical field and machine learning will be described. Next section will cover a brief introduction to the medical field. Then, in Section 3 some well-known machine learning algorithms are introduced and some motivations for machine learning in the medical field will be presented. Section 4 will show how data analytics can help in the medical field. Section 5 will cover the problem statement and finally we will conclude this chapter in Section 6.

2. Medical field

Almost all of us have a general idea about the medical field as we all went through the health care process. Imagine you feel bad and several disease symptoms appear, the first thing you think of is to visit a hospital and consult a doctor.

So, the main purposes of the medical field are to provide good health care, medical materiel and medicines to obtain patient satisfaction. In order to achieve this several organizations such as hospitals, clinics... are equipped with modern medical resources (equipment) used by medical workforce (doctors, midwives,nurses...).

To have a more detailed idea about the medical field let us first introduce some basic concepts in the medical field to clear up ambiguity.

2.1. Medical health care types

For different health problems, different types of healthcare are required. Some patients need normal care while others need extra care. Healthcare can be categorized into three major healthcare types.

2.1.1. Primary healthcare (PHC)

We can also refer to this type of care as essential health care. It mainly focuses on minor but urgent health problems. This type of care is provided by medical generalists, however, opticians and dentists are also considered as PHC providers.

According to the World Health Organization (WHO), five key elements are needed to achieve PHCs goal and are identified as follows:

- Reducing exclusion and social disparities in health;
- Organizing health services around people's needs and expectations;
- Integrating health into all sectors;
- Pursuing collaborative models of policy dialog;
- Increasing stakeholder participation.

2.1.2. Secondary health care

In general, patients are directed by their primary care providers to the secondary health care (SHC) which provides health services for deeper health problems and includes more acute care. SHC is usually afforded in hospitals and clinics and primary care providers . Medical specialists, such as cardiologists, dermatologists and urologists, contribute in SHC. However, secondary health care can also be provided by mental health specialists such as psychiatrists and occupational therapists.

2.1.3. Tertiary health care

This type provides highly specialized care for inpatients which includes advanced medical investigation and treatment that cannot be provided by primary or secondary health care. Tertiary health care can provide services that include cardiac surgery, cancer management and more...

2.2. Medical health care process

Four main phases can be identified in health care process: Prevention, Detection, Diagnostic and Treatment (see Figs. 1–4).

Prevention- Everyone knows the famous proverb " *prevention is better than cure*", therefore it is the first step of the healthcare process which principle objective is to keep patients healthy. This phase includes, for example, physical fitness, good diet and clean drinking water to prevent overweight, bed nets against malaria, or no tobacco to avoid lung cancer.

Detection- Both doctors and public must be conscious of their own health status, they should regularly control their measure hypertension or blood sugar and rapidly consult a doctor when unusual health problems appear.

Diagnostics- When unusual symptoms are detected diagnostics are needed. In this phase, information must be gathered, integrated and carefully interpreted with the aim of detecting a working treatment. Diagnosis can be done in several ways, the patient can be interviewed, a physical exam can be conducted, diagnostic testing can be performed, or doctors can discuss to clear up vagueness. An accurate diagnosis could ensure that a patient's health problem could be correctly understood, hence, appropriate decisions will be made to achieve satisfying health outcome.

Treatment- An effective treatment for each diagnosis is given in order to provide, coordinate, or control patient's health status.

Although the health care process can prevent, detect, diagnose and treat some diseases, it still suffers from several drawbacks.

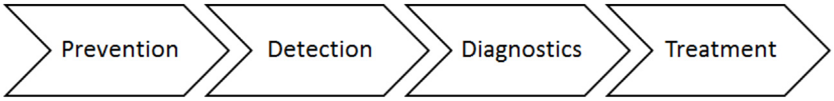

Fig. 1. Health care process.

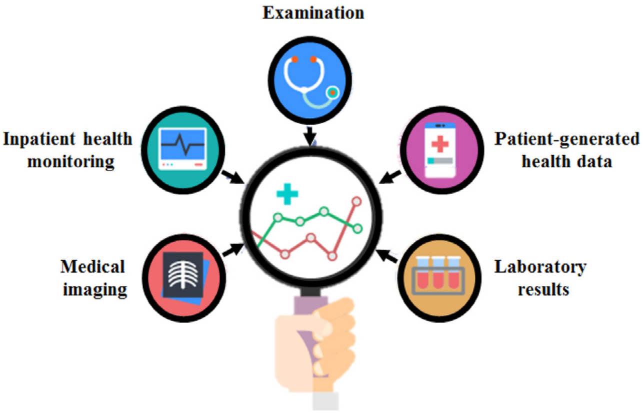

Fig. 2. Medical data sources.

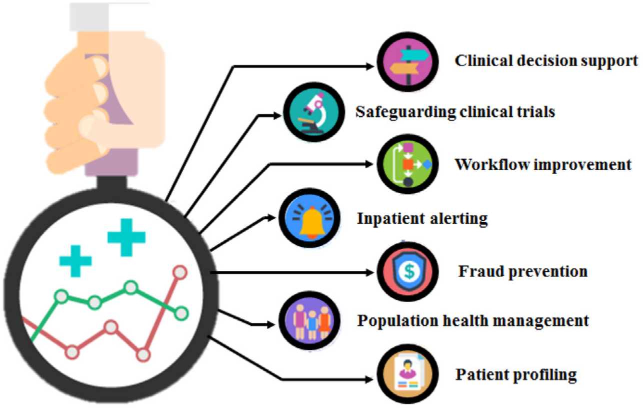
Clinical decision support
Safeguarding clinical trials
Workflow improvement
Inpatient alerting
Fraud prevention
Population health management
Patient profiling

Fig. 3. Motivations for medical data analytics.

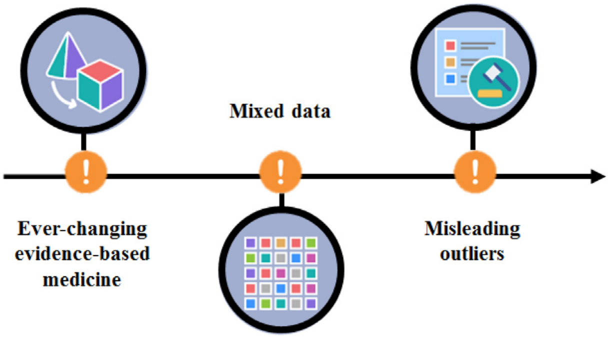

Fig. 4. Medical data analytics challenges.

Difficulties lay in the fact that diagnosis can generate a huge amount of data which makes it complicated and time-consuming to be interpreted. For example x-ray images of cancer tumors in their early stages can be very complicated to be identified and detecting it can take hours. Besides, when we observe at the statistical numbers in the past few years, we can conclude that many new diseases have emerged and others have increased. For example, 19% of Tunisians are diabetic according to the National Institute of Nutrition and Food Technology. According to the Regional Directorate of Health, 2200 new cases of breast cancer are detected each year, which means that almost 30% of Tunisian women are affected by this disease. Besides, in 2015, cancer was the second death cause with about 8.8 million deaths all over the world according to the World Health Organization.

Doctors are working hard to improve the healthcare process, but it can still be improved to show better results and save more human lives. Hence, Machine Learning emerged to facilitate and improve the medical field in several ways.

3. Machine learning

In the last decades, machine learning has become very popular and various machine learning methods have been developed. This section will overview basic concepts of machine learning, in addition to the most commonly used machine learning algorithms.

Machine learning is a field of study of artificial intelligence (AI) which allows machines to be more intelligent without human intervention , i.e., it gives machines the ability to learn by themselves using previous observations and experiences. It focuses on designing, analyzing, developing and implementing methods that can access data and use it to learn by themselves. Therefore, Observations such as examples, direct experience or instructions are needed for better decision-making.

3.1. Machine learning algorithms

Each machine learning algorithm uses a well-defined learning technique that best suits the algorithm's objective. The four main learning methods are supervised, unsupervised, semi-supervised and reinforcement learning, and each methods follows a different learning strategy.

3.1.1. Supervised learning

Supervised learning's name comes from the fact that these kinds of methods need a training set for learning, which is very similar to a teacher supervising the learning process. Such methods are used when possible outputs and correct answers for labeled training sets are already given. This type of learning can be used to reduce classification and regression complexity. **Classification** Classification is a method which aims to identify to which group of subclasses an object belongs. For example, in order to learn a classification algorithm how to correctly classify animals we must provide a training set which contains properly-labeled images of all different animal species and some extra characteristics for training. Naive Bayes Classifier, the Support Vector Machine and artificial neural networks can be used to effectively solve classification problems.

3.1.2. Unsupervised learning

Another technique for learning is unsupervised learning which is closer to true artificial intelligence. With unsupervised algorithms a machine is able to identify complex processes without guidance, i.e., it begins to learn from unlabeled data. In this case, the same animal classification algorithm of the previous supervised learning example will compare all images, separate them into groups based on some similarities and assign to each group his own label. Unsupervised learning can be used for clustering, association analysis and dimensionality reduction.

Clustering Clustering is a fundamental data mining task used for unsupervised learning. It is the process of separating data into meaningful subclasses, called clusters, based on the similarity between data objects, i.e. the likeliness of data objects is a major criteria for clustering.

3.1.3. *Semi-supervised learning*

The name of this learning technique comes from the fact that both labeled and unlabeled data are used in the learning process to overcome supervised and unsupervised methods' drawbacks. This technique needs a supervised learning algorithm to be trained on a labeled training set, then an unsupervised learning algorithm must be applied in order to generate new labeled examples which are added to the old labeled training set for supervised learning.

3.1.4. *Reinforcement learning*

This machine learning method allows machines or agents to learn their ideal behavior, how to act, in a specific situation based on previous experience. Machines collect their training set ("this action was good, this action was bad") in order to improve performance.

Table 1 represents a brief overview of major pros and cons of the discussed machine learning algorithms.

3.2. *Machine learning in medical field*

Today, Machine learning is widely adapted in various domains such as telecommunication, finance, search engines, medical field ... In medical field, machine learning can facilitate very complicated and time-consuming tasks. Thus it is applied in various medical sectors.

3.2.1. *Disease identification/diagnosis*

Disease identification is one major motivation from which the medical field can largely benefit. Machine learning can be applied in the medical field to help physicians save time by detecting diseases in their early stages. Cancer detection is a widely-studied area in machine learning research. In 2010, Microsoft started a machine learning-based project called "InnerEye" [9] which can detect brain tumors and identify its stage in minutes, while this task can take hours when performed by humans. Other researches focused on breast cancer detection [10] in order to facilitate analyzing breast cancer diagnoses and detecting the disease at an early stage.

3.2.2. *Drug discovery/manufacturing*

Other health projects were launched by IBM and Google by applying machine learning in order to discover drug in its early days. These projects outstrips the limits of drug discovery by reducing the drug fabrication process costs and time, with traditional drug discovery it takes years just to develop one new drug. Microsoft worked on "Project Hanover" [11] that uses machine learning to personalize cancer treatment. Previous cancer researches are used to help oncologists create a personalized cancer treatment plan for all different types of cancer.

3.2.3. *Robot surgery*

Machine learning even covers the surgery sector, where it can help identifying different body parts or even really perform surgery. For instance, the "Da Vinci" robot [9] was created to perform robot surgery, Da Vinci can be used in cases with fine details and in tight spaces which can barely done by human themselves.

3.2.4. *Medical data analytics*

Other Machine Learning methods are applied for medical data analytics, analytics can be very useful and interesting to better comprehend the causes of disease evolution, predict diseases and even prevent diseases. Instead of spending hours or even days on interpreting data, applying "ML" could help physicians save patients' lives by saving time.

Many other sectors exist in which machine learning is effectively used and motivations for machine learning will be never-ending. In our further work we will only concentrate on machine learning for medical data analytics as it will concern our further work.

4. Medical data analytics

Data analytics, in general, is the science of examining raw data with the purpose of interpreting and drawing conclusions on the found information. Data can be examined by the traditional usage of queries or by using new methods such as applying machine learning methods on data. In this section we will discover where different medical data comes from. Then we will discuss main motivations for medical data analytics.

4.1. *Medical data sources*

Huge amounts of data are generated everyday in the medical field from which healthcare can benefit when it is properly analyzed. Medical data analytics is capable of improving the health-care quality by reducing the costs of treatment, predicting diseases in order to avoid their evolution and improving prevention methods.

Where does all the information come from?

To obtain accurate medical data analysis, we should know where all the ever-growing data comes from. In healthcare data comes from different sources such as examination, patient-generated health data, laboratory results, medical imaging and inpatient health monitoring.

Examination- Examination generates data that was gathered by a doctor while examining the patient's condition. Examination data includes personal data (sex, age, insurance, contact...), health data (symptoms, childhood diseases, allergies, ...) and transaction data (billing records) if available.

patient-generated health data (PGHD)- Family members, smart medical devices (such as smart watches and health apps) or even the patient can provide patient-generated health data. PGHD contains objective data, which includes heart rate, body temperature, blood pressure ..., and subjective data such as the patient's mood, sleep and more. For instance, PGHD can be used to enhance the chronic disease management quality.

Laboratory results- From laboratory results other data can be collected which is especially valuable for diagnostics and treatment decision making. Test descriptions include fluid or tissue data (blood, urine...) and other characteristics (volume, methods used for the test, time stamp...)

Medical imaging- Another type of data can be gathered from medical imaging, which contains images in 2D and 3D formats. Data can be collected from radiology (MRI, mammography...), nuclear medicine and optical imaging.

Inpatient health monitoring- In this case, continuous data is generated, i.e. streaming data, like cardiac status, pulse rate, pulse oximetry and more. This information can help different care areas like the emergency care, anesthesia and many others.

Table 1
Pros and cons of machine learning algorithms.

Algorithm	Pros	Cons
Naive Bayes Classifier [1]	+ Easy to implement + Computationally fast + Handles high dimensions	− Assumption dependent − Sensitive to outliers
SVM [1]	+ Can handle linearly separable and non-linearly separable data + Handles high dimensions	− Sensitive to outliers
Artificial Neural Network [2]	+ Fast learning + Flexible	− Difficult to implement − Sensitive to outliers
Linear Regression [3]	+ Easy to implement + Good performance for linearly separable data	− Cannot handle non-linearly separable data − Sensitive to outliers
Logistic Regression [4]	+ Robust	− Large sample size − Sensitive to outliers
Decision Tree [5]	+ Easy to understand + Low computational time	− Complex (for large trees) − Loss of innovation (only past experience can go into the tree branching) − Sensitive to outliers
K-means [6]	+ Easy to implement + Computationally fast + Easy to interpret	− K difficult to select − Sensitive to different densities − Sensitive to outliers
A priori [7]	+ Easy to implement	− High computational time − Sensitive to outliers
Principle components analysis [8]	+ Can handle high-dimensional data	− Sensitive to outliers

4.2. Why analyzing medical data?

It is necessary that all these enormous amounts of medical data are properly analyzed. But why are analysis so important? In this section some major motivations for medical data analytics will be discussed.

Clinical decision support- Medical data analytics can improve the clinical decision-making of evidence-based medicine and diagnosis support. Evidence-based medicine is the result of combining extracted observations from medical data with previous experience in similar cases. It is used to discover the best working treatment for a patient and can predict and avoid any possible disease evolution. Diagnosis support uses described symptoms, test results and other data in order to suggest procedures to confirm a disease.

Safeguarding clinical trials- Analyzing existing clinical trials can help improve future trials in suggesting best-working treatment and reducing trial failures.

Workflow improvement- Data analytics may improve the care activity quality by evaluating performance, understanding clinical processes and identifying major activity obstacles.

Inpatient alerting- For some care processes as surgery, post-surgery recovery and rehabilitation data must be continuously analyzed in order to alert positive or negative trends or any health problem. In these cases, data analytics ensure that the care process passes successfully.

Fraud prevention- Data analytics may be used to prevent improper billing or fraudulent activities, therefore transaction data and billing records must be properly analyzed.

Population health management- Population health management may be very important for positive health outcomes, hence,it can be improved by analyzing diagnoses and chronic diseases. Here, data analytics can detect disease outbreaks.

Patient profiling- Advanced medical data analytics, such as predictive analysis, can be applied to identify patients with high health risks.

5. Machine learning applications in healthcare

Thanks to latest technology progress, great masses of medical data are obtained. These large data holds precious information. Thus, clustering is needed for analyzing this amount of medical data. It has been proven effective for Knowledge discovery from large data. However, It is used to cluster large amounts of medical image data of the human brain to identify structures of interest.

One of the most know applications of cluster analysis in medical domain is neuroimaging. And as mentioned in [12] Advanced Magnetic Resonance Imaging(MRI) techniques allow unprecedented insights into the complicated processes in the brain. As well as, it was employed for clinical studies where an enormous amount of data needs to be processed to discover patterns describing the structure and function of the healthy brain and its alternations related with diseases. Some clustering methods are employed for neuroimaging applications such as segmentation of fiber tracks and lesions, also methods that may deal with multimodal imaging data. In fact, clustering is an effective tool that might help for diagnosis analysis purposes as well.

One of the principle activities in neuroscience is image segmentation to split the voxels into homogeneous sub-partitions that correspond to tissue types. Volume Pixel or short voxels define a value on a regular image grid in three-dimensional space. Voxels that does not belong to any shared tissue type can be labeled as pathological structures. As, information about alike neuroimages may lead to advanced diagnosis. In fact, for image partitioning there are also other familiar tools as graph partitioning-based, model-based, and region growing-based approaches, which may be used to puzzle out this task effectively. In comparison with clustering, most of them require user interaction, e.g., defining a seed point or labeled training data in order to realize brain parcellation. Another asset of clustering compared to other tools is that no training data are required. This is particularly significant in clinical studies where we have a restricted number of patients with infrequent pathological conditions and want to make use of most of the data.

In this section we will take a look on the tow main applications in neuroscience which are image segmentation and Pattern Discovery in fMRI:

5.1. Image segmentation

Image segmentation refers partitioning an image into mutually exclusive areas. In medical research studies, determining structures of interest may be considered as one of the most critical tasks. Yet, segmentation still a difficult mission due to varying image contents with various intensity values, image noise, non-uniform object-of-interest etc. Many algorithms and techniques for image segmentation are available but there is still a requirement to design effective, quick tools for medical image segmentation ranging from fiber tracking to lesion detection.

5.1.1. Fiber tracking

Fiber Tracking [13] is a helpful tool, e.g., for procedure planning when the surgical extraction takes place of tumor tissue near to main neural pathways. Particularly, this is applied to tumors close to the visual, speech and motor cortex. In order to predict in advance how to get into the tumor area in the least invasive way, the fibers usually get rendered before the actual procedure. This is realized by first pre-processing the image including registration and motion compensation. Usually, a great number of individual fibers get extracted but the most important task is to find out significant groups of fibers from the Diffusion Weighted Images (DWIs). In practice, a specialist is needed in order to filter out the excess fiber tracks. To perform this task, they mark a specific region of interest (ROI) to detect fibers that passes via this target area by integrating their domain knowledge. The hole process seems to be time consuming and also biased, notably for various patients. This truth constitutes the principle motivation for employing clustering approaches to handle the problem of grouping individual fiber tracks more effectively and more precisely. Density-based clustering methods [14] are suitable for extracting meaningful fibers in a robust fashion. They are able to identify outlier and discover arbitrary shaped clusters. In Two recent clustering algorithm have been developed which are based on an adaptive DBSCAN approach for fiber tracking. In [15] an adapted dynamic time warping (DTW) was designed to define fiber similarity in order to catch local similarity between fibers belonging to the same bundle but having dissimilar start and end points. In paper [16], Warped Longest Common Sub-sequence approach (WLCS) was developed in order to measure similarity between fibers.

5.1.2. Lesion detection

Brain lesion [17] is one of pathological structures that appears in multiple sclerosis, Glioma and Stroke(MS), and may drive to serious neurological damage. One major feature of such lesions is that they most often show hyper-intense (shiny regions in FLAIR Figure), or T1 Gadolinium MR images and hypo-intense (dark regions in T1- or T2-weighted images). The brain include three principle tissue types: white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). Identifying such tissues is achieved by T1-weighted images or similar MRI modalities. It is useful to combine different images from various modalities while clustering tasks in order to obtain a better overview of physical condition in the brain.

Some clustering algorithms are used to attempt to detect the healthy tissue compartments such as WM, GM, and CSF to execute image segmentation process, in order to discover lesions. Knowing the distribution of healthy tissue compartments is often critical to distinguish them from the pathological ones. Pathological voxels can form a significant cluster themselves or just show up as identical outliers, based on the lesion type. In order to detect the parameters of the intensity distributions of WM, GM, CSF, most methods use tissue atlases that represent an average segmentation among a certain number of patients that is then transformed into a probabilistic map for the various tissue types.

5.1.3. Pattern discovery in fMRI images

Pattern discovery in fMRI [18] (Functional MRI, is the localization of neuronal activity in the brain that is triggered by stimuli such as, e.g., finger movement or visual inputs) refers to addressing the problem of clustering fMRI time series in groups of voxels with similar activation. All fMRI data are high-dimensional in nature were each voxel refers to a number of signals in the time range. For displaying the structure of interest that is often composed of many voxels of a particular brain region clustering approaches are used. In similar cases, it is needed to apply brain parcellations in order to divide the brain into non-overlapping regions.

There are several ways to perform this task in literature. But the most efficient parcellation methods are mixture models, variants of k-means or hierarchical clustering. They clustered interaction patterns to discover feature patterns in patients suffering from pain disorder which unlike healthy controls.

6. Problem statement

Although medical data analytics have improved the medical field quality it still faces several challenges. These challenges must be carefully studied in order to obtain effective solutions.

6.1. Ever-changing evidence-based medicine

The first main problem consists in the fact that evidence-based medicine constantly changes with time. Data analysis will always lead to new observations and better conclusions which may exclude previous knowledge, therefore the previous evidence must be updated. For instance, the American Cancer Society continuously updates screening recommendations for early breast cancer detection. Their latest guidelines are released as follows:

> *"The latest guideline applies to women at average risk for breast cancer. Among other recommendations, it says all women should begin having yearly mammograms by age 45, and can change to having mammograms every other year beginning at age 55. Women should have the choice to start screening with yearly mammograms as early as age 40 if they want to".- American Cancer Society.*

6.2. Mixed data

Another challenge emerges when medical data analytics faces unequal data; i.e. data comes from different data sources and are not in the same format. Data can be grouped based on different criteria, the following data types are most commonly retrieved in data analytics.

- **Univariate Data:** for this data type analysis are made on one variable.
- **Multivariate Data:** for this data type more than two variables per observation are used in analysis.
- **High Dimensional Data:** When the number of dimensions, i.e. number of attributes are very high we talk about this type of data.

Analytics approaches that might seem effective for one data type, may not work for other data. Therefore, it is very important to consider data types when analyzing data.

6.3. Misleading outliers

In the medical field data analysis should be performed very carefully as it is related to the patients' health outcome. A little mistake could lead to wrong decision-making, wrong treatments and hence negative health outcome.

Medical data analytics' precision could be affected by misleading or wrong data, also known as noises or outliers. In fact, outliers are data instances that extremely deviate from well defined norms of a data set or given concepts of expected behavior. Therefore, the presence of outliers in medical data has become a major challenge, as outliers could really mislead all analysis. These outliers can be caused by measurement errors, malfunctioning equipment, inherent data variability or human mistake.

For example, Wisconsin Breast cancer (Original) Dataset[1] contains missing values i.e. unrecorded attributes which can be caused by human mistakes. In such case, information is useless and mislead analysis, to obtain better analysis performance they must be deleted. However, these outliers not always carry errors or useless information. For instance, an outlier in Wisconsin Breast cancer (Original) Data set can indicate that a patient may suffer from a malignant tumor. In this case, outliers must be kept for further study and positive health outcome.

Hence, outliers must be carefully detected and studied in order to improve not only medical data analytics results but also decision-making towards outliers.

Are outliers useful or not?

Should outliers be removed or not?

In order to this, we must first know how to properly identify outlying data, i.e., based on which criteria a data point is called an outlier. Some researches used distances between different data objects while others the density of data, other criteria exist and will be discussed in further work. The process of identifying outlying data points is known as outlier detection. Outlier detection has been studied in the last decades, covering several domains such as fraud detection and network anomaly detection. In these domains the detection of outliers showed considerable improvements. Applying outlier detection in the medical field will be a very complicated but also a very important task. It can help us obtain better data analytics' performance, save time and thus save human lives.

7. Conclusion

Machine learning algorithms are widely adapted in the medical field for their incredible ability of fast learning. They have shown revolutionary improvements in several medical care areas especially in medical data analytics. However, these methods still suffer from some challenges, ever-changing evidence-based medicine, mixed data and misleading outliers.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.

[2] F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain., Psychol. Rev. 65 (6) (1958) 386.

[3] K.H. Zou, K. Tuncali, S.G. Silverman, Correlation and simple linear regression, Radiology 227 (3) (2003) 617–628.

[4] D.R. Cox, The regression analysis of binary sequences, J. R. Stat. Soc. Ser. B Stat. Methodol. (1958) 215–242.

[5] J.R. Quinlan, Induction of decision trees, Mach. Learn. 1 (1) (1986) 81–106.

[6] J. MacQueen, et al., Some methods for classification and analysis of multivariate observations, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, Oakland, CA, USA, 1967, pp. 281–297.

[7] R. Agrawal, R. Srikant, et al., Fast algorithms for mining association rules, in: Proc. 20th Int. Conf. Very Large Data Bases, VLDB, Vol. 1215, 1994, pp. 487–499.

[8] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemom. Intell. Lab. Syst. 2 (1–3) (1987) 37–52.

[9] D. Faggella, Machine learning healthcare applications - 2018 and beyond, 2018, https://www.techemergence.com/machine-learning-healthcare-applications/, accessed: 2018-04-03.

[10] C.-H. Chen, A hybrid intelligent model of analyzing clinical breast cancer data using clustering techniques with feature selection, Appl. Soft Comput. 20 (2014) 4–14.

[11] A. Linn, How microsoft computer scientists and researchers are working to 'solve' cancer, 2018, https://news.microsoft.com/stories/computingcancer/, accessed: 2018-04-10.

[12] M.J. Barkovich, Y. Li, R.S. Desikan, A.J. Barkovich, D. Xu, Challenges in pediatric neuroimaging, NeuroImage (2018) http://dx.doi.org/10.1016/j.neuroimage.2018.04.044, URL http://www.sciencedirect.com/science/article/pii/S1053811918303549.

[13] M. Tobin, Lhcb upgrade: Scintillating fibre tracker, Nucl. Instrum. Methods Phys. Res. A 824 (2016) 148–151, http://dx.doi.org/10.1016/j.nima.2015.10.100, Frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors. URL http://www.sciencedirect.com/science/article/pii/S016890021501339X.

[14] A. Smiti, Z. Elouedi, Dbscan-gm: An improved clustering method based on gaussian means and dbscan techniques, in: International Conference on Intelligent Engineering Systems (INES), IEEE Computer Society, 2012, pp. 573–578.

[15] J. Shao, K. Hahn, Q. Yang, C. Böhm, A.M. Wohlschläger, N. Myers, C. Plant, Combining time series similarity with density-based clustering to identify fiber bundles in the human brain, in: ICDM Workshops, IEEE Computer Society, 2010, pp. 747–754.

[16] S.T. Mai, S. Goebl, C. Plant, A similarity model and segmentation algorithm for white matter fiber tracts, in: ICDM, IEEE Computer Society, 2012, pp. 1014–1019.

[17] B. Chen, L. Wang, J. Sun, H. Chen, Y. Fu, S. Lan, Y. Huang, Z. Xu, Diverse lesion detection from retinal images by subspace learning over normal samples, Neurocomputing 297 (2018) 59–70, http://dx.doi.org/10.1016/j.neucom.2018.03.023, URL http://www.sciencedirect.com/science/article/pii/S0925231218303096.

[18] R.T. Thibault, A. MacPherson, M. Lifshitz, R.R. Roth, A. Raz, Neurofeedback with fmri: A critical systematic review, NeuroImage 172 (2018) 786–807, http://dx.doi.org/10.1016/j.neuroimage.2017.12.071, URL http://www.sciencedirect.com/science/article/pii/S1053811917310959.

[1] WDBC dataset is from UCI ML Repository: http://archive.ics.uci.edu/ml.