2020 11th International Conference on Information and Communication Systems (ICICS)

The Use of Big Data Analytics in Medical Images: a Survey

1st Mohammed Shatnawi
Computer Information Systems Department
Jordan University of Science and Technology
Irbid, Jordan
mshatnawi@just.edu.jo

2nd Muneer Bani Yassein
Computer Science Department
Jordan University of Science and Technology
Irbid, Jordan
masadeh@just.edu.jo

3rd Ibtihal Jalabneh
Computer Information Systems Department
Jordan University of Science and Technology
Irbid, Jordan
isaljalabneh18@cit.just.edu.jo

2020 11th International Conference on Information and Communication Systems (ICICS) 978-1-7281-6227-0/20/$31.00 ©2020 IEEE 10.1109/ICICS49469.2020.239514

***Abstract*— as data grows there is a needs for the technology that has the capacity to deal with this growth. As data grows in all fields on a daily basis, the challenge becomes harder in finding the right technological tools that can handle this expansion. This growth poses new challenges in terms of creating the technology or solutions that can deal with this growth or even the challenges arisen on the organizational level. Many sectors have already adopted big data. However, in the area of healthcare sector, it is still in the early stages of adoption. The problem is not in the data itself but the lack of information that can be gained with the help of technologies like Hadoop to generate this information. Hadoop also helps in generating meaningful real time analysis which can predict emergency situations or help in planning, strategizing, and decision making but its uses in the ECG field is still limited. The data generated by healthcare organizations can also have more and higher value if we can elicit more useful information from it using proper data analytics. By adopting big data analytics, organizations can have the chance to manage and analyze volumes of massive data more efficiently. A comparison of the potential between machine learning and big data was discussed and finally the available literature on the use of big data and deep learning in the ECG filed and its possible applications.**

Keywords*— *Big data analytics, healthcare analytics, analytics, big data, Hadoop

I. Introduction

Technological advancements and innovations are always useful in enhancing different sectors, including the health sector.

The importance of big data gains attention now more than ever before, as it's considered as the engine for socio- economic growth [2]. Big Data can be generated through different ways such as instant transactions, educational records, hospital records, mobile devices, social media, and online shopping over diverse activities across different subjects.

The origin of the term big data appeared out of the increased size of all types of data on the internet, whether that is a structured or unstructured data. All different sources and types of data still require analyzing and managing and therefore the amnesty and complexity of the big data requires new ways of management and innovative tools of storing, processing, mapping and analyzing which cannot be dealt with using the conventional data processing tools [1].

The strategies used to achieve success with big data in healthcare analytics are: implementing big data governance, developing an information-sharing culture, training key personnel to use big data analytics, incorporating cloud computing into the organizations, as well as generating new business ideas from big data analytics. To conclude, big data analytics could be an effective IT artifact to create IT capabilities and business benefits potentially and how health organizations can leverage big data analytics as means to create business value for health care [3].

Big data analytics is a vital function of big data as we can explore more within the data and link up matched patterns and interrelationships amongst different items and individuals' interests, for example, through connecting some links from a huge data set [1]

Big data is different from conventional sources of data in the main four dimensions or ways. First, the volume as it has more voluminous, it can be argued for example, that data that was generated within two days in 2011 was much bigger than the data generated since civilizations inception. The form of this data can take trillions of gigabytes. However, it can also be measured in petabytes and zettabytes [1]. Second, Big data means or represents velocity; for example, there are many generated data that come from streams online or tools that can report instant data related to a particular subject. Third, the variety of data types and sources, which can be massive in most cases, therefore, requiring analysis and management. These data can be in different forms and formats such as data collected from Video content or communicative messages or posts on social media or even wearable devices in which all need processing, storing, analysis and management. Fourth, veracity, as it can be abnormal and noisy, which again, needs sorting out and analyzing [4] [1]

Big data as a concept is not something we did not know before; however, the way it's defined is being shaped constantly. The four dimensions of speed, size, type, and complexity can reflect and categories what big data can be like; therefore, this requires constant research on creating new hardware and software mechanisms to make better use of storing, analyzing and visualizing the data. In the healthcare sector, the dimensions of big data are embodied within the daily practices that the healthcare sector generates on a daily basis [8]. The daily examples of medical practices can be seen in many places such as in the laboratories and pharmacies or written notes from physicians or prescriptions or administrative data or even medical imaging in the ECH scans (electrocardiogram), which is the main subject of this paper. An electrocardiogram (ECG) is a medical examination that recognizes heart problems by measuring the electrical movement produced by the heart as it contracts. If the ECG shows a different shape, it could suggest a heart intricacy.

In using Big Data, it will be possible to reach more and better personalized, targeted, and robust diagnostic paths. In this context so far, Machine or Deep Learning can contribute significantly or give the chance to produce radical innovation in the medical field and especially cardiac diseases. Until recently, the research on physiological signals, including the ECG, could count on mathematical or statistical approaches. The use of big data analytics can help in prevention, prediction, management and treatment of cardiovascular diseases [6].

A. Big data analytics and machine learning applications and performance

When looking at the difference between Big Data analytics and Machine Learning, one can say that big data is the process in which a large volume of data gets generated and collected.

Afterward, the data can be analyzed to uncover hidden patterns within the data that cannot be seen otherwise, such as customer choices, business trends that can help institutions make informed and better decision making. Whereas, machine learning is a subset field within Artificial Intelligence in which a machine or computer gets educated on how to perform a complex task that humans naturally do not know how to do so or achieve that easily. Machine learning is advancing so quickly in our daily lives, and there is a large amount of machine learning activities happening around us these days. These activities can be seen through Google's recommendations of suitable content based on previous search or history of the users or other activities such as The Uber application and the way it minimizes the waiting time once a customer makes the order.

One can say that there is no direct relation between big data and machine learning; however, big data can be used in machine learning. Hadoop is an open-source software used to store data that can afterward be processed. Whereas, Machine learning is data analysis technology. However, Hadoop is more than a database. It's a software that can provide an ecosystem that gives the users to store but, at the same time, do computing on the data.

B. A comparison of the general foundations between Machine Learning and Big Data.

Table 1 below summarizes the major differences between the big data and machine learning technologies.

TABLE I. DEFERRENCES BETWEEN MACHINE LEARNING AND BIG DATA

Feature	Machine learning	Big Data
Learning difference	Teaching the machine to teach itself through learning from existing data available	Whereas, Big data analytics picks or study from existing data to look for common themes or patterns that aid in predictions or decision-making lifecycle
Application	Machine learning can be seen in advanced technology around us today such as Google's or social media recommendations engines or self-flying jets	Big Data analytics can be used for different purposes through collecting the data and make research on the data to discover patterns that can help in understanding consumer behavior or trends or other services such as in healthcare and finances and others
Task	Machine learning takes the algorithms and educate it to automatically do a task or learn how to do it from the collected data	Big Data analytics can only uncover patterns through classifications and sequence analysis of the data
Size	Machine learning learns how to train algorithms from different sized data, However, the more data you have, the more advanced models one can develop.	Big Data analytics deals and stores large and huge volumes of datasets
Main Goal	Machine learning learns from a trained data and predicts or estimates future outcomes	Big Data analytics works as a storage for the data and spot out patterns in the data that can help in research and decision-making process

II. RELATED WORK

Big data adoption came alongside the huge amount of data that is generated by the internet of things (IoT), Big data analytics tools were needed to create and make more use and sense out of the data generated which can help in understanding users' behaviors or patterns of behaviors when using things on the internet [11].

Most big data applications today use the state-of-the-art technologies such as Hadoop and Spark. These applications can be suitable in fields such as business analytics in which it can deal with the voluminous aspect of the data through attempting to mine information that can be relevant in certain contexts. However, these technologies still struggle to deal with the humongous heterogeneous data environment where a massive number of events can happen frequently and at the same time, which can be grouped under the dimensions of velocity and variety together happening at the same time [5].

As alluded to earlier, big data essentially entails the concept of complexity, too many, too much, too fast and too big. It's important to stress that the traditional technologies and systems such as HTTP, XML, JSON, Web Services and RESTful applications are not enough and limited to handle such complex environment. This poses the urgency for new solutions and technologies that can work suitably around the whole range and dimensions of big data. Hadoop solutions and technologies emphasize the volume dimension and are not adequate for fast data, however, for the time being Hadoop can still be a good match for big data and is capable of processing very large data sets in parallel [5].

There are various benefits and advantages gained from using big data analytics of which mainly can be seen through the reduction of costs, providing new services, and enhanced decision making [5]. Big data gives organizations an opportunity to manage their data and be more competitive around other competitors as it improves efficiency and productivity and caters for customers more effectively [18], [19] [21].

A. *Big Data in health care and medicine*

The use of big data analytics in the industry of health can entail many benefits, such as the use of data to detect diseases at a very early stage based on previous data gathered and processed. Also, it can provide more customized treatment and help in analysing the data collected about patients who had suffered before from a particular disease or symptoms and therefore understanding some conditions better and thus providing better treatments [7].

The authors of [7] have explained the lifecycle of big data in five cycles starting from first data collection by storing all the data needed, such as patients with the same symptoms or medical images in HDFS. Second, data cleaning through processing the data detecting the errors from the records and removing it or fixing it. Third, classification of the data by sorting the relevant data and putting them together in classifications to make sense out of the data. Fourth, data modeling or analysis, which essentially involves the analyzing the data that had been already classified.

Similarly, the authors in [3] have discussed the benefits and capabilities of big data analytics for health organizations using the quantitative approach, which comprises various analytical techniques such as the descriptive one. With this strategy, the new database allows data being transferred between traditional and new operating systems to support health organizations to meet the patients' needs and future market trends, thus improve financial performance. Furthermore, Health organizations like hospitals will also benefit from the big data operations by having the issues of data storage, cost, and size of enterprises solved. In addition, their research demonstrates the IT values in business and how it is related to the importance of analytical personnel, which is in the future, is going to get involved in the big data analytics framework.

The volume of health data is on the rise, and big data analytics is expected to help in transforming the way providers of healthcare use technologies by gaining better understanding and insight from all parties involved from clinical to other repositories to make informed decisions which will have a reflection on the whole different range of services provided by health organizations [22].

In medicine, the increased electronic health records alongside more developments in computer capabilities have resulted producing data at an unprecedented scale, which resulted in an increased interest in the use of big data analytics in this area [12]. Big data analytics have helped in providing and generating new knowledge in a much faster way than traditional scientific approaches. It helped in bringing more understanding to data gathered in biology and pathophysiology. There is also more promise that Big data analytics can bring to medicine such as more personalized and precision medicine for patients meaning more advanced field and better accuracy in treating patients. Similarly, it can also help in earlier diagnosis [12].

Big data in health and medicine come from various sources of structured and unstructured data. The structured data has a high level of organization which makes it easy to deal with, such as clinical laboratory tests. The unstructured data comes through different sources such as written notes or laboratory reports and medical imagining [12]. Therefore, this made big data analytics one of the most important scientific discussions lately. This paper presents a brief overview of big data analytics in medical imaging with considering the importance of contemporary machine learning or Hadoop as the technology that can deal with data generated from medical images.

As revolutionizing big data is in different fields, it can be said that it can play a significant role in all healthcare practices leading to an improvement in disease exploration and care delivery. It can also be noted that the use of technology in general in the medical field has been on therise.

Computer-aided technology, or what is also called computer-aided diagnosis, has become an essential part of the clinical work in the detection of different diseases [9].

Medical imaging is simply creating and producing visual representations of the body from the inside. These images can help in making a clinical analysis of the tissues, diagnosis of the bones and internal organs of the body, and their functions.

The huge capabilities and potential of computer-aided technologies in the healthcare field, especially in medical imagining, can pose the urgency to improve and develop more in this area of research. Thus, the data produced by these technologies can have different features and dimensionalities. Big data analytics is argued to be a solution in handling and processing this level of data which can help doctors or radiologists or healthcare providers, in general, in which can then provide even better services than ever before as this means they can store the data and at the same time they have the right tools to process this data and to also do research on it. Research in this respect can vary depending on the body part these images are taken from, and therefore researchers have a more specialized area to research. However, in this paper, I'm presenting in the general benefits of using big data analytics in general [13].

Many initial stages of produced algorithms had provided some solutions when the data was small to moderate-sized; however, they could not cope or eventually failed as the scale of the data went up [10]. Currently, the scale on the available data produced in the market is massive. In this respect, we currently know is that there are no available tools in the market to cope with the different dimensionalities the data produce as every single tool has its strengths and limitations at the same time. Therefore, there needs more exploration and research on finding the best algorithmic tools that can be used to cope with particularly targeted dimensionalities depending on the nature of data produced.

There are six types of medical images data that are produced by different forms of radiation with a tissue. These types of medical images can vary from X-ray to more sophisticated images produced by functional magnetic resonance imagining, which can provide images with complex spatial and temporal resolutions. The first type is "radiography," which simply represents the forms of 2D X-rays. The technology of X-rays was invented long before the modern computers by Wilhelm Konrad Rongtgen, who discovered that radiation could travel through different subjects and be captured on a photographic plate [14].

The second type is "X-ray Computed Tomography," which is a development or next stage development of the use of modern computers with radiography. This technology brought new great advancements in the X-ray technology comprising of producing thousands of images that can be produced in 2D or 3D image formats [15].

The third type is "Magnetic Resonance Imaging (MRI)," this type uses a powerful magnet to produce images that cannot be captured using X-rays or CT, for example, ligaments, tendons, joints, and cartilage. The MRI data is captured at a high complex rate of multiple slices per second. This high speed, however, is compromised, producing low spatial resolution [16].

The fourth type is Functional MR Imaging (fMRI), which is a sophisticated technology that produces images that maps the functional connectivity in the brain by locating and measuring the wave of the fluctuations in the blood-oxygen- level [17].

The fifth type is "Nuclear Medicine Imaging" this type of imaging is produced using nuclear radiation, which is used to diagnose tumours, infections, and thyroid or bone scintigraphy. The last type is Ultrasound, which is a method that produces real-time images through a non-invasive method [13]

These different types of medical images create high volumes and a massive spectrum of different image acquisition methodologies for varied purposes. These technologies provide high-resolution data, and which results in an exponential volume of data that demands advanced analytical methods and computing with high performance.

B. *The use of Deep learning and ESG scans and signals and their applications*

The authors in [24] has emphasized the quality of the deep learning which performs better for several databases to develop the model by using all the available input rather than the classical analysis and machine classification methods. Their work differentiates between the physiological signals as a data source and the manual physiological signals. The first is more effective than the second one in which human errors could occur because of human fatigue. Therefore, the problems of the decision-making process willdisappear.

Regarding arrhythmia detection, first of all, arrhythmias are irregularities in the rhythm of the heartbeat which may happen to anyone, and usually, a Holter device is used to record ECG data for the long term. Thus, the task of having a recognition of abnormal heartbeats through the use of deep learning is important. In this regard, the work of [27] has provided an ECG beat classification through providing a deep learning approach on Arrhythmia MIT-BIH database which achieved staggering results in terms of accuracy, sensitivity and specificity.

Regarding the work on machine learning in the field of ECG signal, first, deep learning is a representation-based learning that essentially includes an input layer, then hidden layers, and an output layer. This is considered a systematic procedure or a network that can be fed with data and learns the required representations for classification [20]. The convolutional neural networks (CNN) is one of the most popular neural network techniques. The work in [25], has developed a classification methodology through the use of deep learning in particular deep convolutional neural networks (CNN). The methodology focused on a life-threatening cardiac arrhythmia which is the condition of "Paroxysmal atrial fibrillation". They used a massive volume of raw ECG time-series data in their approach. The benefits of their approach are to reduce or potentially remove the need to using the human element expertise as well as the process of extraction in accordance to the different cardiac arrhythmias.

Machine learning can be argued to provide many solutions in analysing and classifying ECG data. One of the solutions is the application of deep learning architectures where first layers of convolutional neurons act as point extractors and where layers at the end are used for obtaining final decision about ECG classes [25].

Similarly, the use of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. The authors of [23] highlighted how the electrocardiogram helps in the detection of some heart diseases like myocardial infarction (MI) by proposing an approach to automatically detect the MI by ECG signals. Also, it has become possible to detect even the unknown heartbeats that are without noise automatically. Their work proposed a CNN algorithm that is introduced to clinical settings to aid the clinicians. The ECG signals used to diagnose heart attacks and abnormalities but with small amplitude and duration, and to overcome this problem they used the CAD system "computer aided diagnosis due to its fast, objective and reliable analysis. There are also other techniques employed by researchers for the automated detection of MI, the one performed better was using the EGC without noise. The authors' work in [23] concluded that the early diagnosis for heart function prevents attacks and damages to have early treatment.

Unlike other studies, this paper talks about how the ECG databases belonging to different stages of CHF and CNN could anticipate detecting different cardiac diseases like dilated, ischemic, hypertrophic cardiomyopathy. As long as the CNN model is well trained it can be introduced in the healthcare industries as an adjunct tool to assist cardiologists in providing quick and reliable second opinions in the diagnosis. In this study they find the manual ECG signals in CHF and traditional techniques are inadequate, also to find distinctive features between normal and CHF signals difficult and involves a lot of time and effort. As a result, they decided to propose deep learning in order to optimize the performance of an automated CHF diagnosis system. It is an approach where the network learns and picks up distinct characteristics automatically based on the input ECG signals [24] [26].

In 2019, a new paper has been published on the use of ECG signals in biometric recognition systems. Although ECG based biometric systems are usually less accurate than technologies based on other physiological traits, but they have been successfully used to perform biometric recognition in wide range of applications. To evaluate the performance of deep ECG with samples acquired in scenarios different from the ones used to train the CNN (ambulatory at rest vs Holter during real life) obtaining better identification accuracy or comparable to the best performing methods in the literature. In addition to a widely used validation strategy has been used to achieving accuracy in line with the best performing methods in the literature. All results are particularly relevant and proves the applicability of deep ECG for ECG signals collected in heterogeneous application scenarios [28]

III. CONCLUSION

Many sectors have already adopted big data however, research is still limited in the area of healthcare, medicine, medical imaging and ECG sectors, and it is still in the early stages of adoption. Hadoop can be argued to have the potential to help in generating meaningful real time analysis which can predict emergency situations or help in planning, strategizing, and decision making. The data generated by healthcare organizations can also have more and higher value if we can elicit more useful information from it using proper data analytics. This survey highlighted the potential of the use big data analytics in all sectors and in the health care sector in particular. This paper discussed the importance of big data analytics in medicine, medical imaging and ECG stressing the potential of its use in improving the ECG filed. A comparison of the potential between machine learning and big data was discussed and finally the available literature on the use of big data and deep learning in the ECG filed and its possible applications.

REFERENCES

[1] Abdel-Basset, M., Mohamed, M., Smarandache, F., & Chang, V. (2018). Neutrosophic association rule mining algorithm for big data analysis. Symmetry, 10(4), 106.

[2] Tsou, M. H. (2015). Research challenges and opportunities in mapping social media and Big Data. Cartography and Geographic Information Science, 42(sup1), 70-74.

[3] Tu, C., He, X., Shuai, Z., & Jiang, F. (2017). Big data issues in smart grid–A review. Renewable and Sustainable Energy Reviews, 79, 1099- 1107.

[4] Aguilar, S. J. (2018). Learning analytics: At the nexus of big data, digital innovation, and social justice in education. TechTrends, 62(1), 37-45.

[5] Mahmood, Z. (Ed.). (2016). Data science and big data computing: frameworks and methodologies. Springer.

[6] Ur Rehman, M. H., Yaqoob, I., Salah, K., Imran, M., Jayaraman, P. P., & Perera, C. (2019). The role of big data analytics in industrial Internet of Things. Future Generation Computer Systems, 99, 247-259.

[7] Archenaa, J., & Anita, E. M. (2015). A survey of big data analytics in healthcare and government. Procedia Computer Science, 50, 408-413.

[8] Bian, J., Topaloglu, U., & Yu, F. (2012, October). Towards large-scale twitter mining for drug-related adverse events. In Proceedings of the 2012 international workshop on Smart health and wellbeing (pp. 25-32). ACM.

[9] Doi, K. (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Computerized medical imaging and graphics, 31(4-5), 198-211.

[10] Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. National science review, 1(2), 293-314.

[11] Riggins, F. J., & Wamba, S. F. (2015, January). Research directions on the adoption, usage, and impact of the internet of things through the use of big data analytics. In 2015 48th Hawaii International Conference on System Sciences (pp. 1531-1540). IEEE.

[12] McCue, M. E., & McCoy, A. M. (2017). The scope of big data in onemedicine: unprecedented opportunities and challenges. Frontiers in veterinary science, 4, 194.

[13] Tahmassebi, A., Ehtemami, A., Mohebali, B., Gandomi, A. H., Pinker, K., & Meyer-Baese, A. (2019, May). Big data analytics in medical imaging using deep learning. In Big Data: Learning, Analytics, and Applications (Vol. 10989, p. 109890E). International Society for Optics and Photonics.

[14] Webb, A., & Kagadis, G. C. (2003). Introduction to biomedical imaging. Medical Physics, 30(8), 2267-2267.

[15] Momose, A., Takeda, T., Itai, Y., & Hirano, K. (1996). Phase–contrast X–ray computed tomography for observing biological soft tissues. Nature medicine, 2(4), 473.

[16] Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. proceedings of the National Academy of Sciences, 87(24), 9868-9872.

[17] Ehtemami, A. (2016). Statistical data analysis of resting state fMRI: A study of nicotine addiction treatment (Doctoral dissertation, Ph. D. Dissertation. The Florida State University).

[18] Yoder, K. K. (2013). Basic pet data analysis techniques. In Positron Emission Tomography-Recent Developments in Instrumentation, Research and Clinical Oncological Practice. IntechOpen.

[19] Cipresso, C., Rundo, F., Conoci, S., & Parenti, R. Big Data in Preclinical ECG Alterations Research. Biomedical Journal, 1, 2.

[20] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436.

[21] Acharya, U. R., Fujita, H., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adam,

[22] M. (2017). Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. Information Sciences, 415, 190-198.

[23] Acharya, U. R., Fujita, H., Oh, S. L., Hagiwara, Y., Tan, J. H., Adam, M., & San Tan, R. (2017). Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals. Applied Intelligence, 49(1), 16-27.

[24] Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S., & Acharya, U. R. (2018). Deep learning for healthcare applications based on physiological signals: A review. Computer methods and programs in biomedicine, 161, 1-13.

[25] Pourbabaee, B., Roshtkhari, M. J., & Khorasani, K. (2017). Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 48(12), 2095-2104.

[26] Al Rahhal, M. M., Bazi, Y., AlHichri, H., Alajlan, N., Melgani, F., & Yager, R. R. (2016). Deep learning approach for active classification of electrocardiogram signals. Information Sciences, 345, 340-354.

[27] Sannino, G., & De Pietro, G. (2018). A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. Future Generation Computer Systems, 86, 446-455.

[28] Varatharajan, R., Manogaran, G., & Priyan, M. K. (2018). A big data classification approach using LDA with an enhanced SVM method for ECG signals in cloud computing. Multimedia Tools and Applications, 77(8), 10195-10215.