www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 7 July 2017, Page No. 21963-21967
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i7.11

Applications of Big Data Analytics and Machine Learning Techniques in Health Care Sectors

Mrs.J.Sukanya M.C.A., M.Phil[1], S. Vijaya Kumar M.Sc., M.Phil.,[2]
[1]Assistant Professor, Department of Computer Science
M.V.Muthiah Government Arts College(W), Dindigul-624001
Email:sukanrajam76@gmail.com
[2]Research Scholar
M.S.University,Tirunelveli
Email:ssvjaykumar@gmail.com

Abstract: ***Big Data in Healthcare refers to handling the large and complex sets of data and making it accessible at finger tips. Earlier, the data in healthcare used to reach a level of complexity from where it used to become next to impossible to extract that data and bring it to use. This leads to slow growth in healthcare industry. Big Data becomes much more interesting when it is used in healthcare analytics. The other potential benefits that big data provides is the detection (Analysis) of diseases beforehand which allows doctors to act treat proactively. This will also help in detecting fraud in healthcare more quickly and efficiently, By bringing into use of statistical tools and algorithms, there could be a faster development of more accurately targeted vaccines. The tools are used to help developing the cost effective ways for discovering more clinically relevant ways to analyse the disease and treat the patients in an effective manner. The objective of this paper analyzes importance of big data and the various steps involved in machine learning techniques in healthcare. This paper also identifies the big data analytics are helping to realize the goals of diagnosing, treating, helping, and healing all patients in need of healthcare.***

1. Introduction

Big data platform is capable of processing terabytes and petabytes of data, as a result of which, data analysis becomes easier. Big data analytics has offered a new way to healthcare organizations to develop actionable insights, organize their future vision, boost up the outcomes and reduce time to value. This approach is also helpful to provide insightful information to the healthcare enterprises regarding their management, planning and the measurements. Environmental factors influencing the physical and psychological health include air, water and soil. The recent approach conceptualizes the environment more broadly to encompass a range of human-made physical and social features that are affected by public policy. In layman's terms, environmental health is the health impact of the air breathe, the water we drink, the homes we live in, the soil growing the food we eat and the many other environmental exposures in our lives. The study of EH is not new. As early as 400 BC, Hippocrates said that one's health depends on the air one breathe, the water one drinks and the environment in which one lives.[1]

The Machine Learning technique for Healthcare in developing algorithms that is used to identify the complex patterns with large amount of data. This technique implies the ways to make intelligent data-driven decisions. It focus on developing and applying machine learning and data mining tools to an array of different challenging problems from clinical genomic analysis, through designing clinical decision support systems[2].There are two general categories of algorithms: unsupervised and supervised. Unsupervised machine-learning algorithms are typically used to group large amounts of data. Unsupervised algorithms can be used to generate hypotheses, and thus, often precede use of a supervised algorithm. Supervised machine-learning algorithms start out with a hypothesis and categories that are set out in advance. These results are then used to make predictions based on out-of-sample data for which the outcome of interest is not known.

2. Related Work

Health is determined by several factors including genetic inheritance, personal behaviors, access to quality health care, and the general external environment (such as the quality of air, water, and housing conditions). In addition, a growing body of research has documented associations between social and cultural factors and health (Berkman and Kawachi, 2000; Marmot and Wilkinson, 2006)-[3] The influence of social and cultural variables on health involves dimensions of both *time* (critical stages in the life course and the effects of cumulative exposure) as well as place (multiple levels of exposure). Big data is created, stored, and disseminated through traditional and mobile Internet, smartphones, smart TV, sensor- and RFID-based ubiquitous networks, and social media [4]

Given a set of clinical cases that act as examples, learning in intelligent systems can be achieved using ML

methods that are able to produce a systematic description of those clinical features that unique indicate those conditions. As **John Smith, Senior Manager for Intelligent information Systems at IBM Research**, says one of the most promising near-term applications is for its use in detecting melanoma**.** By feeding a computer with many images of the cancer**,** IBM is planning to teach the system how to recognize features associated with the disease as well as support the physician with text-based medical records concerning diagnosis and treatment protocols.[5]

Disease detection and surveillance systems provide epidemiologic intelligence that allows health officials to deploy preventive measures and help clinic and hospital administrators make optimal staffing and stocking decisions.[12]

3. Characteristics of big data in healthcare

Big data is characterized by variety (structure to unstructured data, textual or multimedia, data graphs), velocity (how fast data comes in) and volume (massive amounts of data points, new and historical data), explained Kinson Ho, MSc, a product architect at Agfa Healthcare.

Not all of this is new, the increase in velocity is not outrageous for most of us, and the increase in volume is offset by the increase in storage and processing, Ho said. The real game changer of big data is starting something that's specialized: special hardware with special processing, and now big data technology develops an opportunity to use common hardware

4. Types of big data Analytics

The following four types of big data analytics need to be considered at the time of big data analysis**.**

4.1. Predictive

Predictive analysis identifies past data patterns and provides a list of likely outcomes for a given situation. By studying recent and historical data, predictive analysis presents you with a forecast of what may happen in the future. One common type of predictive analysis is sentiment analysis, in which the model predicts the sentiment score based on data it has. Predictive analysis can also be very useful in optimizing customer relationship management**.**

4.2. Prescriptive

Prescriptive analysis reveals actions that should be taken and provides recommendations for next steps, letting you answer your business questions in a focused manner. Prescriptive analysis can be used to improve drug development, reduce time to market for new medicines and find the right patients for clinical trials.

4.3. Diagnostic

Businesses use this type of analysis to complete root-cause analyses and uncover patterns in their business processes. Ultimately, it can help identify factors that directly or indirectly affect their bottom line. Business growth can often be driven by the smarter decisions made as a result of diagnostic analysis.

4.4. Descriptive

Finally, descriptive analysis examines what is happening in real-time based on incoming data. Descriptive analysis is often referred to as the simplest type, since it allows converting big data into useful bite-sized nuggets [8]

5. Big Data Analytics in Healthcare

Data analytics in healthcare has the potential of transforming the ways healthcare organizations offer their sophisticated clinical facilities. This powerful analysis method is ideal to improve the care, save patient's life and lower the cost of the health facilities charges. Data analytics in healthcare can be used to raise the standards in following fields

Public Health: By analyzing disease patterns and recording disease outbreaks, public health issues can be improved with analytics approach. Large amount of data can be used to determine needs, offer required services and predict and prevent the future crises to benefit the population.

Electronic Medical Record or EMR: An EMR contains the standard (structured and unstructured) medical data that can be evaluated with the data analytic approach to predict patients at risk and provide him effective care.

Patient Profile Analytics: Advanced analytics can be applied to patients' profile for identifying individuals who could benefit from proactive approach. This may include lifestyle changes.

Genomic Analytics: The data analytic approach can be effectively included in genomic analytics to make this approach a part of regular medical care decision process.

Fraud Deduction: This data analytics approach helps analyze greater number of claim requests to curtail down fraud cases. An effective analysis can help reduce fraud, waste and abuse.

Safety Monitoring: Data analytics can also be used to analyze real time large volumes of brisk data in hospitals. The approach may help in the safety monitoring and negative event prediction.[6]

MRI and PET (Positron emission tomography): it can provide assessment of both cerebral metabolism and amyloid. Computer algorithms are being adapted for medical purposes to identify specific elements of images in order to offer early detection of disease or flag up the warning signs to physicians.

Video streaming Statistics: It is used for the purposes of understanding the consumption patterns (behavioural and optimizing viewing experience).

Patient care in radiology: Big data in radiology is more about decision support than anything else and plays an

important role in defining the way radiologists use clinical decision support systems to assist them in reading images.

Pediatric cardiology: analytics are applied to make patient-specific recommendations for treatment. The Pediatric Cardiac Critical Care Consortium (PC4) uses big data to try and improve the quality of care by collecting data on clinical practice and outcomes from each patient's medical record and analyzing the data to provide clinicians with timely performance feedback.[7]

Episode analytics. The move to value-based payments is well underway and accelerating. The shift is putting unprecedented pressure on health care providers to better manage the cost and quality of care they deliver.

Population health analytics. The promise of big data and analytics in managing population health is one of the most hyped yet least understood opportunities in health care today.

Virtual care and wearable health care technologies. Technology is helping providers drive virtual care initiatives to increase quality of care and provide patients with more access, but there is the question: How secure is the ecosystem in which more and more personal health information is being exposed?

Project Data Sphere. SAS is playing a key role in the development of industry-wide pharmaceutical data transparency. The goal is a secure, globally accessible data and analysis environment where multiple organizations can share anonymized clinical trial information. This would help scientists learn from research more quickly, and thus speed improvements in care.[11]

Disease progression models- Most recently disease progression was studied on the organ or system level. Using vitals, labs and invasive procedures, the patient disease state is observed as the function of the organ or the system [18]

6. Machine Learning Techniques in Healthcare

Machine learning (ML) provides methods, techniques and tools that can help solving diagnostic and prognostics problems in a variety of medical domains. ML is being used for the analysis of the importance of clinical parameters and of their combinations for prognosis, eg. prediction of disease progression, for the extraction of medical knowledge for outcomes research, for therapy planning and support, and for overall patient management. [9]

6.1 Types of Machine Learning Algorithms

Four different types of machine learning algorithms are available that can be organized into taxonomy based on the desired outcome of the algorithm or the type of input available for training the machine. Thompson noted, "The terminology used in machine learning is different than that used for statistics. For example, in machine learning, a target is called a label, while in statistics it's called a dependent variable." [17]

The key types of machine learning include:

- ✓ Supervised learning.
- ✓ Unsupervised learning.
- ✓ Semi supervised learning.
- ✓ Reinforcement learning.

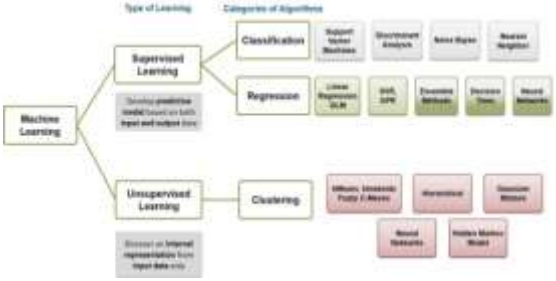

Figure 1. Types of Machine Learning

Supervised learning is a type of machine learning that uses a known dataset (called the training dataset) to make predictions. The training dataset includes input data and labelled response values. Supervised machine learning techniques are more suitable for medical data classification [16].Unsupervised learning is a type of machine learning used to draw inferences from datasets consisting of input data without labelled responses.

6.2 Applications of Machine Learning Techniques in Health Care

Machine learning algorithms are effective in recognizing complex patterns within rich and massive data. This capability is particularly well-suited to medical applications, especially those that depend on complex proteomic and genomic measurements. As a result, machine learning is frequently used in various disease diagnosis and detection. In clinical applications machine learning algorithms can produce better decisions about treatment plans for patients by means of providing effective healthcare system

6.2.1 Discrete Event Simulation-Health care organizations are using this technique to predict wait times for patients in emergency department waiting rooms. The models use factors such as staffing levels, patient data, emergency department charts, and even the layout of the emergency room itself to predict wait times

6.2.2 Free-text physician notes-IBM researchers have found a way to extract heart failure diagnosis criteria from free-text physician notes method. They developed a machine learning algorithm that combs through physicians free-form text notes (in the electronic health records) and synthesize the text using a technique called "Natural Language Processing" (NLP). Similar to the way a cardiologist can read through another physician's notes and figure out whether a patient has heart failure, computers can now do the same.

6.2.3 Predicting Strokes and Seizures-Singapore-based start-up Healint launched an app called JustShakeIt

that enables a user to send an emergency alert to emergency contacts and/or caregivers simply by shaking the phone with one hand. The program uses a machine learning algorithm to distinguish between actual emergency shakes and everyday jostling. In addition to the JustShakeIt app, Healint is working on a model that analyzes patients' cell phone accelerometer data to help identify warning signs for chronic neurological conditions

6.2.4 Proprietary predictive model-Using this predictive model, hospitals can predict emergency room admissions. Thus the application of machine learning may benefit patients either by reducing costs, improving accuracy, or disseminating expertise that is in short supply

6.2.5 Machine Learning Techniques in Numerous Disease Predictions and Diagnosis: Machine learning plays a key role in many radiology applications. Machine learning identifies complex patterns automatically and helps radiologists make intelligent decisions on radiology data such as conventional radiographs, CT, MRI, and PET images and radiology reports [19].

7. Steps to apply Machine Learning in Health Care Data

1. **Define the Problem-**Describe the problem informally and formally and list assumptions and similar problems. List the technique for solving the problem, the benefits a solution provides and how the solution will be used with health care system.
2. **Select Data and Prepare a model-**Data preparation with a data analysis phase that involves summarizing the attributes and visualizing them using scatter plots and histograms.
   Step 1: Data Selection: Consider what data is available, what data is missing and what data can be removed.
   Step 2: Data Pre-processing: Organize your selected data by formatting, cleaning and sampling from it.
   Step 3: Data Transformation: Transform pre-processed data ready for machine learning by engineering features using scaling, attribute decomposition and attribute aggregation.[14]
3. **Spot Check Algorithms –** Test and validate the process. Spot checking algorithms is a part of the process of applied machine learning. On a new problem, we need to quickly determine which type or class of algorithms is good at picking out the structure in the problem and which are not. Loading up a bunch of standard machine learning algorithms into the data set harness and performing a formal experiment. There are three key benefits of spot-checking algorithms in machine learning problems are speed, objective and result.[15]
4. **Improve Result-** If better performance is needed, it becomes necessary to utilize more advanced strategies to augment the performance of the model .Using Machine Learning techniques it will reduce the variance of the performance measure.
   The process of improving results involves:
   - ✓ **Algorithm Tuning**: where discovering the best models are treated like a search problem through model parameter space.
   - ✓ **Ensemble Methods**: where the predictions made by multiple models are combined.
   - ✓ **Extreme Feature Engineering**: where the attribute decomposition and aggregation seen in data preparation is pushed to the limits.
5. **Apply Result [13]** -Depending on the type of problem to solve, the presentation of results will be very different. There are two main facets to making use of the results of your machine learning endeavour:
   - ✓ Report the results
   - ✓ Operationalize the system

After these steps have been completed, if the model appears to be performing satisfactorily, it can be deployed for its intended task. The model may be utilized to provide score data for predict the disease, for projections of Electronic Medical Record, to generate useful insight for decision making or research, or to automate tasks. Machine learning is closely related to computational statistics, a procedure which focuses in prediction through the use of computers. ML methods are implemented with optimization techniques, which deliver methods, theory and application domains to the field.

8. Conclusion

With more and more data available, machine learning techniques are becoming increasingly popular as they get better at looking at massive amounts of data. The most important challenges in clinical practice and biomedical research include the need to develop and apply novel tools for the effective integration, analysis and interpretation of complex biomedical data with the aim to identify testable hypothesis, and build accurate models. Big data Analytics gives a great boost to leverage the benefits of chaotic environment in healthcare. Using these new techniques it is easier to develop therapics and products. By comparing the effectiveness of machine learning models are used in health care delivery and services and assessing health. Cloud computing model is a perfect match for big data with healthcare since cloud computing provides unlimited resource on demand. This allows aggregation of multiple disparate workloads with varying performance goals into very large clusters. Incorporating big data in healthcare clearly has the ability to transform the industries. Personalized medicines and early monitoring and diagnosis are also expected due to the appropriate analysis of big data.

9. References

[1]Hippocrates On airs,waters, and places. Retrived April, 21,2010
[2]https://www.research.ibm.com/haifa/dept/vst/analytics.shtml
[3] https://www.ncbi.nlm.nih.gov/books/NBK19924/

[4] Song TM. Efficient utilization of big data on health and welfare. Health Welf Policy Forum. 2012;(193):68–76.
[5]http://www.technologyreview.com/news/540141/why-ibm-just bought-billions-of-medical-images-for-watson-to-look-at/
[6]http://www.builtinla.com/blog/significant-benefits-big-data-analytics-healthcare-industry
[7] http://www.scimage.com/the-future-of-big-data/
[8]http://www.smartshifttech.com/big-data-and-analytics-which-type-analytics-does-your-business-need
[9]http://www.dcs.bbk.ac.uk/~gmagoulas/ACAI99_workshop.pdf
[10]http://medcitynews.com/2016/06/machine-learning-algorithms-and-healthcare-litigation/diagram
[11]http://www.information-management.com/news/big-data-analytics/how-big-data-keep-united-healthcare-nimble-10026746-1.html?
[12] Brownstein, J. S. & Mandl, *K. D.* Reengineering real time outbreak detection systems for influenza epidemic monitoring. Am Med Inform Assoc, Annual Symposium Proceedings vol. 2006, p. 866 (2006).
[13]http://machinelearningmastery.com/process-for-working-through-machine-learning-problems/
[14]http://machinelearningmastery.com/how-to-prepare-data-for-machine-learning/
[15]http://machinelearningmastery.com/why-you-should-be-spot-checking-algorithms-on-your-machine-learning-problems/
[16]S. Anto, Dr.S.Chandramathi , Supervised Machine Learning Approaches for Medical Data Set Classification - A Rev IJCST Vol. 2, Iss ue 4, Oct . - Dec. 2011
[17]http://www.sas.com/content/dam/SAS/en_us/doc/conclusionpaper1/statistics-machine-learning-at-scale-107284.pdf
[18]https://www.nhlbi.nih.gov/health/health topics /topics /copd
[19] Shijun Wang, Ronald M. Summers, "Machine Learning and Radiology", Med Image Anal., 2012.