Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 3, March 2017, pg.20 – 27

A Study on Data Analytics: Internet of Things & Health-Care

N.Nalini, P.Suvithavani

AP/CSE, Sri Shakthi Institute of Engineering and Technology

AP/CSE, Sri Shakthi Institute of Engineering and Technology

Abstract- Farming is the foundation of India. Nowadays individuals in India losing the trust in cultivating the land because of inclement weather. The Agriculture joins its hands with technology to improve the productivity of agriculture. The improvements are in the field of selection of seed, Usage of fertilizer to avoid the soil infertility, tillage equipment, decision on effective usage of resources. These upgrades are accomplished by the IoT Analytics on the sensor gathered information. The smart cultivating helps an awesome degree Agriculture Sustainability. Health care analytics is a data-rich domain. Nowadays health care data is voluminous and thus become intricate. The health care industries are focused on Value-based business rather than volume based business. This results in increasing demand for big data analytics in health care. Big Data can afford insight to take the right decisions at the right time for the patients. This requires analytics techniques to facilitate the medical practice with right care and high quality.

Keywords: Analytics, AgicultureIoT, IoT, Healthcare Analytics, Big data Analytics, Sources of data.

I. Big Data Management in Health care

The 4 Vs of big data relevant to health care[2] are volume, variety, veracity and value. In healthcare big data refers to electronic health data sets which are so large and complex. A database is created that takes patient names and addresses from one system and matches it up with scheduled appointments from another system or integrating claims data with clinical notes from the EHR[Electronic Health Record]. Mending multiple sources of information together into a centralized databank accessed by reporting or a query system can provide a more in-depth and actionable snapshot of history of the patient, diagnoses, treatments, socio-economic challenges and other risk profiles. It is also used to envisage epidemics, cure disease, improve quality of life and avoid preventable deaths. Hospitals and healthcare systems are excellent depot of big data (like patient records, test reports, medical images etc.) that can be utilized to lower the cost in healthcare and to get better reliability and efficiency. The trend in healthcare is changing from cure to prevention.

II. Sources of health care system

The familiar data types in public health Research are quantitative data and qualitative data. Quantitative data is quantifiable and used for comparisons of data. It entails behaviour of people, counting the number of people, their conditions and discrete events. The examples of quantitative data are temperature, weight, age, total number of people suffering from the particular disease. Qualitative data consists of non-numerical data. The words can be used to portray

the health related issues. The qualitative data cannot be measured, but can be observed. The examples of qualitative data are Female/Male, Non-smoker/Smoker etc.

2.1 Diverse kinds of Data Sources:

The kinds of data in health science include clinical data and scientific data. The clinical data focus on data's pertaining to patient or health care such as health surveys, epidemiological data. The scientific data includes data based on the bench sciences. All data's are recorded based on ICD (International Classification of Diseases). It is a terminology of signs, symptoms, diseases, and procedure codes maintained by the World Health Organization(WHO).

The data sources are broadly classified into Primary data sources and secondary data sources. The primary data analysis means the individual person or squad of researcher's designs, collects, analyses the data**.** This gathered data can be used to answer the research questions. The individual or the squad can control the data collection process. The advantages of primary data are guaranteed data quality, lower the number of missing values and measure the consistency of the instruments. The secondary sources depend on the existing data. Existing data means the data already collected for another purpose. This existing data is used for answering the research questions. The advantages of the secondary data sources are low cost, time to collect data is less, and huge data samples can be obtained.

2.1.1 Electronic Health Records(EHR) Data:

The significant source of health care analytics comes from electronic health care records[EHR].Electronic health records(EHR) is the digitized patient record. It makes the information accessible wherever and whenever needed. It contains each and every details of patient's health in one place. The EHR data's can either be structured data or unstructured data.

2.1.1.1 Structured & Unstructured EHR Data:

Structured data is prepared by health care professional by filling the data's in a specific format. All the data's are captured and classified in a database. It is reliable and dwell in predefined fields within the record. This makes it easy to observe, share and right to use.

Unstructured data is Vague. It is characterized by the static pages of health information. The data's are not organized and irregular. The types of unstructured data are emails, word processing files, PDF files, spreadsheets, Digital images, video, audio, social media posts. Most of the information is converted into structured data. This is quite complex process. This incurs high cost and very time consuming.

2.1.2 Clinical Text Mining:

The medical records can be structured, unstructured and may contain textual records. Text mining is the process of extracting high-quality information from the unstructured data. Some of the text mining techniques include sentiment analysis, Categorization, information extraction. Text mining in electronic medical records[EMR] has finer benefits. Clinical text mining[6]is used to discover the unknown disease, greater patient stratification, better targeting of medicines, and unknown side effects of the drugs. The Natural language Processing[NLP]can be used for information extraction in clinical text mining. Many mining tools are used for analysing the textual records in health domain. One of the examples of clinical text mining tool is clinical Text Analysis and Knowledge Extraction System (cTAKES).

2.1.3. Medical Imaging Data:

The unstructured image data includes CT scans and X-rays. One of the systems used to store Medical imaging data is PACS(Picture Archival &Communication Systems)[6]. This system is used for storage and retrieval of the images. In medical Retrieval system all the images are stored in biomedical image database. The image data occupies more space and it is complex. This system is divided into two phases training and testing phase. The feature is extracted by applying various algorithms during the training phase. In testing phase the query result is obtained based on the input query image.

2.1.4. Genomic data:

Genomic data is a discipline in genetics. It deals with DNA to assemble, structure and sequence the function of genes. All the data's are collected, stored and processed with the help of software. The Genome database(GDB) is the repository of Human Genome. Genome-wide association studies (GWAS)[6] is used to find out the common genetic factors which influence health and disease.

2.1.5 Behavioural data:

It comes from social network data and mobility sensor data. There is some social networks that can be used to track diseases. An online data sharing named as 'Patients like me[6] 'was started in the year 2006 and it is used to follow the other patients those who have the same symptoms and the treatment they have undergone.

III. Health care data Quality

The three terms often used in health care analytics are Health care data, Health care information, Health Care Knowledge. The processed health care data[1] is health care information which in turn processed health care information is health care knowledge. Health care data is the main source of health care information. A health care organization need to have high-quality health care data to obtain high-quality health care information. Some of the causes of Poor health care Data quality are unclear data definitions, unclear data collection guidelines, low interface design, errors in programming, partial data source, inappropriate data format in the source, Data dictionary is lacking or not available,

Data dictionary does not stick on to guidelines, deficient of sufficient data checks, No proper system to rectify detected data errors, No control over adherence to guidelines and data definitions, Illegible handwriting in data source, errors in typing, errors in calculation.

The characteristics of Data quality[1] are Accessibility, Accuracy, Consistency, Comprehensiveness, Currency, Definition, Granularity, Relevancy, Precision and Timeliness. Information technology has remarkable potential as a tool for getting better health care data quality.

A data quality management tool has been published by American Health Information Management Association (AHIMA).The data quality management model focuses[1] on four criteria. They are Application, Collection, warehousing, Analysis. The application describe about the purpose for which the data is collected. Collection portray about the process by which the data elements are accumulated. Warehousing depict the Processes and systems used to archive data and data journals. Analysis is the process of translating data into information which is utilizable for the respective application.

3.1 Health care information standards

The data can be recognized as high quality if it conforms to a standard. A set of "essential principles of healthcare documentation", has been published by The Medical Records Institute (MRI).

Legal health record(LHR)[1]contains the documented services regarding health care which is delivered by a healthcare provider organization and is provided to an individual. It consists of data which is required to identify the patients such as diagnostic images, photographs, tracings, and monitoring strips, as a part of the LHR.

IV. Health care & Real Time

Sensing data's can provide more data in real time. eHealth is most significant due to health monitoring of chronic illnesses, lifesaving in emergency situations. It also has the potential to provide round the clock healthcare to rural and disadvantaged areas. The overall eHealth monitoring framework consists of the following components. Situational awareness sensors, Communication networks, Medical data processing servers, clinic terminals.

Wireless Body Area Networks (WBANs) are the main source of remote and in-hospital health monitoring. It can revolutionize the health and real-time body monitoring industry. Nowadays the rapid development of smart phones, sensors, body sensors and wireless communications pave the way for eHealth monitoring. The Wireless body area networks encompass a number of sensors which are rooted in the patient body along with IOT sensors sensing environmental environment. The sensors are integrated through a controller by which it transmits data to the cloud periodically or on demand. The personal server(PS) runs on Smart phone, PDA or home personal computer which collects the medical data and aggregate them.PS serves as an interface between WBAN sensors, users and other data servers. WBANS is configured and managed by PS which includes registration of sensors nodes, initialization, task allocation and specification, setting up secure communication channels between the sensors. WBAN's are mainly used in emergency care.

V. Technologies that support health care information

The big data analysis can be used for decision making in health care with the aid of machine learning algorithms[4]. The different approaches of machine learning algorithms and data mining can afford better cure to disease, build up personalised medicines and even prevent disease or epidemics. Traditional Machine Learning Algorithms works on centralised databases. There is a necessitate to amend the traditional algorithms or crop up with some common hybrid approaches to manage the large dataset in distributed environment.

Apache Hadoop[5] offer Hadoop Distributed File System storage which takes care of distributed storage and fault tolerance. Scalable data analytics platform with in-memory computing can be done using Spark technology. It support open source environment which has high computing power. Spark is designed for explicit applications like machine learning algorithms and natural language processing. Storm technology is used for streaming process. It is an open source launched by twitter in September 2011.It deals with the map reduce concept of Hadoop. It is put into practice in Clojure language to support machine learning environment. HPCC(High Performance Computing Cluster) deals with the analysis of large amount of data by the MapReduce framework .The programming language used here is Enterprise control language (ECL).The major two advantages of HPCC over Hadoop is scalability and speed. The in memory computing for big data SAP devised a tool called HANA. It processes on the block of data with the aid of advanced parallel architectures and algorithms for higher speed.

5.1Tools and Data analytics platform used in Health care System:

The diverse number of tools[3] is used to progress health care data and analytics. This will prop up descriptive, predictive and prescriptive healthcare data analytics.

1. **Advanced Data Visualization (ADV):**

ADV can handle various data types. It varies from erstwhile standards bars and line chart. It is trouble–free to use. It supports analysts to explore the data widely. ADV can produce best results and to disclose clinical hidden patterns in the data.

2. Presto:

Presto is used to analyze enormous amount of data. It is a distributed SQL Query engine. With the help of presto data can be analysed in few seconds or minutes.

3. Hive:

The large amount of data can be handled using hive. Unlike presto it is not used for processing and analysing data quickly. It performs all excel tasks efficiently. Industries use both presto and Hive for preeminent performance. Presto can access data stored on Hive.

4. Vertica:

It is analogous to Presto. It swathe hefty amount of data including hospitals' data and analytics. It is less expensive as it eliminates costly architecture. It encompasses the feature of scalability. Vertica can improve healthcare by lowering operational costs, accelerating medical reports and analysing patients' patterns.

5. Key Performance Indicators (KPI):

It is a approach that use electronic medical record data to identify human practice and inventions. KPI can get better quality of medical health care for patients who are vulnerable to hospital conditions. It is used to denote noteworthy indicators to be monitored and corrected.

6. Online Analytics Processing (OLAP):

OLAP perform statistical calculation in a high speed through multidimensional organised data and amplify data integrity checking, reporting services and quality control. It gives better tracking of medical records and diagnoses. Thus improving health care decision making system.

7. Online Transaction Processing (OLTP):

It is interrelated to OLAP. It is used to process patient registration, documentation of health records in hospitals, different operations of patients care, and review the results.

8. The Hadoop Distributed File System (HDFS):

Healthcare data analytics system is enhanced using HDFS. It divides huge amount of data into smaller one. The miniature data is distributed across other systems. It eliminates data redundancy. It is mainly focused on assisting diagnosis, treatment planning, monitoring patient's signs and fraud detections.

9. Casandra File System (CFS):

CFS is a distributed system similar to HDFS. It is a designated system used to act upon analytic operation with no single point of failure.

10. Map Reducing System:

Map reducing system handles enormous amount of data. It split the chore into subchore and get together its outputs. It facilitates operational calculations to be performed efficiently. It keeps track on each sever when the chore is being performed. The main advantage of map reducing is high-level of parallelism.

11. Complex Event Processing (CEP):

CEP is recently used in healthcare sectors. It monitors the different states of patient. Complex event processing relates and link the events to real time. Thus it improve EMR and HER systems.

12. Text Mining:

In health care systems text mining tools can be a added value for examining clinical records from hospitals. Text mining can recommend treatment plan that can build up a number of standards and protocols. Treatment can be done based on these developed standards.

13. Cloud Computing:

Cloud computing has greater advantage in health care sector. It offers more flexibility by act in response for the dynamic changes and recent medical updates. It adds a great health care value by lowering costs, improving the productivity and data analysis, providing better security. It lowers the sprain due to voluminous data.one such example is Phillips Health suite platform.

14. Mahout:

A mahout is an apache project. It intends to develop applications that can prop up health care data analytics on Hadoop systems.

15. JAQL:

JAQL is a functional query language used to process huge amount of data. It makes possible parallel processing by translating high level queries into low level ones. JAQL works well with map reducing.

16. AVRO:

AVRO make possible data encoding and serialization. It improves data structure by specify data types, meaning and scheme.

VI. IoT Systems

The IoT has the following important factors that make the IoT System more dominant[7]:

1. IoT Device and Data Gathering:

There is a lot of IoT Device available in the market which collects the data and sends across to the data collection center for the analysis. This functionality will do with the help of API (Application Program Interface). The API reduces our effort in developing the IoT Environment and makes the application success. While Developing the API, it should be easy to integrate and scalable.

2. Data Collection Center and Stream Processing

Nowadays the data from the source are very speed. To handle the high-speed data is the challenging issue. In order to solve this problem, the IoT needs a support from the Real-time Stream Processing which will select the samples from the continues streaming data and do the processing on the main memory to give the instant results. During the creation of algorithm for sampling data in the stream needs to concentrate on the most accurate sample for processing.

3. IoT Analytics

The Analytics are performed on the warehouse, which is populated by the Extract, Transform, and Load. Earlier days, the structure Query Language was used for analysis and finding the valuable information from the data. But now a lot of improvements have been done in this field like Amazon Redshift, Google BigQuery, and HP Vertica.

4. Visualization Tools and Dashboard

The final important component is Visualization tool and dashboard. If we have good analytics algorithm but visualization was poor then the effort made in developing the system will be the waste. So, the visualizing the result in a meaningful way is another challenging factor. In some visualization tool will give the feel like doing the analytics by themselves. The system needs to provide the user-friendly tool, rapid result display in streaming of data and insight analysis.

VII. IOT Analytics

The decade ago only people are connected to the internet but now things are connected to the internet. The sensors are embedded in the things, which will help the things to connect with the internet. So, the wireless sensor network is the heart of IoT. The major role of the sensors is sense the event and shares the data. In the market, different types of sensors are available. Based on our application we will be selecting the sensors and attached the things to make the thing to smart thing. The sensor generates the structured and unstructured data, very speedy data, and the huge size of data. Data Analytics guide us in acquiring best decision-making, accurate findings, and optimization in all areas like agriculture, Medical, Monitoring activities and retail industries. The new trend in Data Analytics is IOT Analytics, in which data are gathered from the sensors and perform the real-time analytics on the sensor data. The new IOT Analytics needs to handle the high-speed and high volume data. The IOT Analytics performs three main tasks gathering of knowledge, finding the unknown events which are useful to us and applying the finding to the relevant applications[9].

Figure 1.IoT Analytics.

The sensor generates the continuous stream of data. The analyst needs to perform the analysis on continuous data and find the useful information. Handling the continuous data from the device is the great challenge for the Data scientist. The machine algorithm is the effective way to learn the stream of data and find the new hidden information from the generated data. The new findings will be displayed with the better visualization technique.

VIII. Sources of IOT Data

The Source of the data is based on the type of an IOT application. The IOT data is widely divided into three categories serving as passive data, active data, and dynamic data. The passive data come from the device without active response. The active data are collected from the IOT device which is actively reacting and send the response. The dynamic data are collected and making the self-decision to do a better performance[10].

The data are sourced from the following applications:

1. Industrial Control Systems
   The IoT plays a vital role in the maintenance/failure of the equipment in the industry. The Cortana Analytics, SAP HANA, and IBM Watson are the machine learning which helps to do predictive analysis on industrial control systems.
2. Business Applications

   The business application gets the data from the CRM, ERP or EAM. Based on the received data the analytics identify the positive and negative transactions in past few quarters. Also, it helps us to identify the customer's behavior. It exhibits the future of the business and guides us on our production and investments.
3. Wearables

   The literal meaning of wearables means that is worn by the humans. The wearable devices are embedded with the sensors, it continually senses the reading and act according to the data. The few examples of wearable devices are IntexFitrist Fitness Band Bluetooth 4.0 , Withings Wireless Blood Pressure Monitor, navigation tools.
4. Agriculture Assistant

   The Agriculture is the back bone of our country. There is a lot of IOT analytics are happening in the IOT analytics domain. The data for agri analytics are from the sensors. These results help us to predict the future plant and lead to sustainable agriculture. There is a lot of research happening in the agriculture domain.
5. Open & Web Data

   The open and web data concentrate on the publicly available social network data such as Facebook post, comments, tweets. Based on these data we can infer the current trends and interest among the group of people.
6. Location

   The Location data are collected from the GPS (Global Positioning System).These data are helping in track the vehicle (Supply and chain Management). It also helps in farm Management.

IoT Analytics Domain

The IoT Analytics are focused on the following domain such as

- Forecasting the agriculture production/manufacture
- Machine learning algorithms
- Failure prediction
- Predictive maintenance
- Supply and chain
- Frequent pattern mining.

IX. AgriIoT

India is very popular for agriculture. Currently, we are struggling to make the profit on agriculture. In order to overcome the difficulties, the IoT can extent their wings in the domain of agriculture. For smart farming, the sensor monitors the soil nutrients, moisture levels, weather monitoring, and identification of pests. With these values, we can do sustainable agriculture with no wastage of water and reduce the usage of fertilizer. The farmers are insisted on using the

mobile APP in which they get the analysis results and based on the result, they will take the decision to get profitable agriculture[10].

9.1Architecture of AgriIoT:

There is no standard framework/architecture for IoT. The architecture can be deployed based on the domain where the application is used. The AgriIoT contains six main layers are present in the architecture of IoT shown below [10].

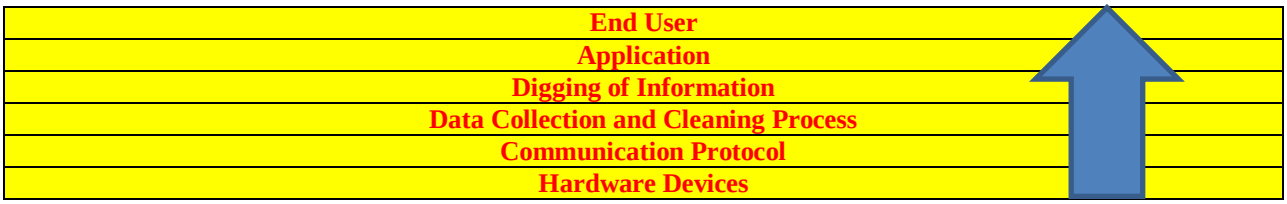

Figure 2.Architecture of AgriIoT

To create the AgriIoT Product first we have choose the appropriate hardware to attach to the problem statements. Then, the data collection and cleaning process will take place. On the collected data information needs to be mined and the application needs to be present the mined results to the end users. The AgriIoT can be applied to the field of Seed Selection, Farm Mechanization, Precision Farming, soil erosion, Agriculture Marketing and Irrigation System.

X. Conclusion and Challenges

The data analytics can get better standards in Public health, Electronic medical records, Patient profile analytics, Gnemoic analytics, Fraud analysis and Safety monitoring. Some of the big data analytic companies that focus on health care domain are Oracle, Sparx IT solutions, IBM, Allscripts Healthcare solutions, Verisk Analytics. Thus analytics in health care is effective for various health care organisation varying from individual general practitioner to multi provider health related organisations. The main challenge in AgriIoT Analytics is to consider the spatial imbalance and temporal imbalance. The analytics will vary based on the location and time. There is a lot of way to collect the information, but selecting the sample from it is a challenge. Because, from the sample only the analytics report is going to generate and decision will take based on report. Since the report is used in real-time processing, the analytics algorithm should generate the report in short duration. The time complexity of an analytics algorithm should minimum. The QoS of the AgriIoT applications are measured with the Precision, Scalability and Security. The AgriIoT not only increase productivity but also helps us do sustainable agriculture.

References:

[1]. karen a. Wager, franceswickham lee , john p. Glaser,"***Health care information systems***",2nd *edition* .989 Market Street, San Francisco:2009,pp. 3-83

[2]. Jasleen Kaur Bains. (2016,4). Big Data Analytics in Healthcare- Its Benefits, Phases and Challenges.[online].6(4),pp-430-435.Available: https://www.ijarcsse.com/docs/papers/Volume_6/4_April2016/V6I4-0143.pdf

[3]. Mohammad Ahmad Alkhatib, Amir Talaei-Khoei, Amir Hossein Ghapanchi. (2015). Analysis of Research in Healthcare Data Analytics. Presented at Australasian Conference on Information Systems .[online].Available: https://acis2015.unisa.edu.au/wp-content/uploads/2015/.../ACIS_2015_paper_107.pdf

[4]. Dharavath Ramesh1, Member, IEEE ,Pranshu Suraj2, and Lokendra Saini3. (2016). Big data Analytics in Healthcare: A SurveyApproach .Presented at International conference on Microelectronics, Computing and Communications (MicroCom).[online].Available: ieeexplore.ieee.org/document/7522520/

[5]. Prof. Jigna Ashish Patel, Dr. Priyanka Sharma. (2014). Big data/or Better Health Planning .Presented at IEEE International Conference on Advances in Engineering & Technology Research (ICAETR - 2014).[online].Available: ieeexplore.ieee.org/document/7012828/

[6]. Jimeng Sun, Chandan K. Reddy. (2013). Big Data Analytics for Healthcare.Presented at SIAM International Conference on Data Mining, Austin.[online].Available: https://www.siam.org/meetings/sdm13/sun.pdf.

[7]. Rajeshwari.D (2015,June), State of the Art of Big Data Analytics: A Survey, International Journal of Computer Applications, . [Online]. pp 39-46. Available: http://research.ijcaonline.org/volume120/number22/pxc3904456.pdf

[8]. Mehdi Mohammadi, Mohammed Aledhar(2015,January),Internet of Things: A Survey on Enabling Technologies, Protocols and Applications, IEEE Communications Surveys & Tutorials, [Online], Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.720.4460&rep=rep1&type=pdf

[9]. Amol Dande ,Ass.Prof. Shaffi K. Pathan (2016, May), Industrial Market Place Survey and Analytics on Internet of Things, International Journal of Advanced Research in Computer and Communication Engineering, [Online], pp 527-529, Available: http://www.ijarcce.com/upload/2016/may-16/IJARCCE%20132.pdf

[10]. ShailajaPatil , Anjali R. Kokate , Dhiraj D. Kadam(2016,August), Precision Agriculture: A Survey, International Journal of Science and Research, [Online], pp 1837-1840, Available: https://www.ijsr.net/archive/v5i8/ART2016967.pdf
[11]. J.M. Barcelo-Ordinas, J.P. Chanet, K.-M. Hou, and J. García-Vidal, "A survey of wireless sensor technologies applied to precision agriculture," in Proc. 9th ECPA conference, Available: https://www.researchgate.net/publication/239937504_A_survey_of_wireless_sensor_technologies_applied_to_precision_agriculture.

